

NASA Technical Memorandum 109925

# NASA STI Program Coordinating Council Twelfth Meeting

April 13, 1994

## Standards

(NASA-TM-109925) NASA STI PROGRAM  
COORDINATING COUNCIL TWELFTH  
MEETING: STANDARDS (NASA) 218 p

N94-36853  
--THRU--  
N94-36860  
Unclass

G3/82 0018958



## ***The NASA STI Program ... in Profile***

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program plays a key part in helping NASA maintain this important role.

The NASA STI Program provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program is also NASA's institutional mechanism for disseminating the results of its research and development activities.

Specialized services that help round out the Program's diverse offerings include creating custom thesauri, translating material to or from 34 foreign languages, building customized databases, organizing and publishing research results ... even providing videos.

**For more information about the NASA STI Program, you can:**

- **Phone** the NASA Access Help Desk at (301) 621-0390
- **Fax** your question to the NASA Access Help Desk at (301) 621-0134
- **E-mail** your question via the Internet to [help@sti.nasa.gov](mailto:help@sti.nasa.gov)
- **Write** to

NASA Access Help Desk  
NASA Center for AeroSpace Information  
800 Elkridge Landing Road  
Linthicum Heights, MD 21090-2934

NASA Technical Memorandum 109925

NASA STI Program  
Coordinating Council  
Twelfth Meeting

April 13, 1994

# Standards



National Aeronautics and Space Administration  
Scientific and Technical Information Program  
Washington, DC 1994

This publication was prepared by the NASA Center for AeroSpace Information,  
800 Elkridge Landing Road, Linthicum Heights, MD 21090-2934, (301) 621-0390.



# TABLE OF CONTENTS

Speakers .....	i, ii
List of Coordinating Council Meetings .....	iii
Document Preparation .....	iv
Welcome and Introductions .....	1
Ms. Patt Sullivan	
STI Architectural Framework Working Group .....	1
Ms. Karen Kaye	
OSI and TCP/IP .....	11
Dr. Lynwood P. Randolph	
A Concise Introduction to MARC .....	53
Mr. Randall K. Barry	
Open System Environments .....	85
Mr. Fritz Schulz	
Open System Environment Procurement .....	116
Mr. Gary Fisher	
Standard Generalized Markup Language .....	159
Mr. Robert Donohue	
Z39.50 and GILS Model .....	161
Mr. Eliot Christian	



**NASA STI PROGRAM  
COORDINATING COUNCIL MEETING**

**STANDARDS**

April 13, 1994  
10:00 a.m. - 4:00 p.m.  
Crystal City Gateway 4  
Conference Room

**Welcome and Introductions**

Ms. Patt Sullivan  
NASA STI Program

**STI Architectural Framework Working Group**

Ms. Karen Kaye  
NASA STI Program

**OSI and TCP/IP**

Dr. Lynwood P. Randolph  
NASA Information Technology Standards Program

**A Concise Introduction to MARC**

Mr. Randall K. Barry  
Library of Congress

**Open System Environments**

Mr. Fritz Schulz  
National Institute of Standards

## **Open System Environment Procurement**

**Mr. Gary Fisher  
National Institute of Standards**

## **Standard Generalized Markup Language**

**Mr. Robert Donohue  
DOE Office of Scientific and Technical Information**

## **Z39.50 and GILS Model**

**Mr. Eliot Christian  
U. S. Geological Survey**

# NASA STI Program Coordinating Council

The NASA Scientific and Technical Information (STI) Program Coordinating Council consists of participants from NASA Headquarters, NASA Centers, and NASA contractors. The Coordinating Council meets periodically to exchange information and pursue topics of vital interest to the NASA STI Program.

## Coordinating Council Meetings

First Meeting	NASA RECON Database	May 23, 1990
Second Meeting	International Acquisition	July 23, 1990
Third Meeting	STI Strategic Plan	November 29, 1990
Fourth Meeting	NACA Documents Database Project	February 7, 1991
Fifth Meeting	Quality	July 1, 1991
Sixth Meeting	Who Are Our Key Users?	October 25, 1991
Seventh Meeting	Acquisitions	January 23, 1992
Eighth Meeting	Using the Internet	June 5, 1992
Ninth Meeting	Total Quality Management	October 28, 1992
Tenth Meeting	Information Retrieval: The Role of Controlled Vocabularies	April 22, 1993
Eleventh Meeting	NASA STI Modernization Plan	September 9, 1993
Twelfth Meeting	Standards	April 13, 1994

## Document Preparation

The following transcription was prepared from the audio tape of the session by the staff of the NASA Center for AeroSpace Information and reviewed by the speakers. The transcription is intended to give the substance of the presentations and does not attempt to exhaustively report comments from the audience. Accompanying viewgraphs immediately follow each presentation.

# Welcome and Introductions

Patt Sullivan welcomed the guests and introduced the first speaker, Karen Kaye.

## STI Architectural Framework Working Group

### Karen Kaye

Thank you, Patt. As you can see on our agenda, we have a pretty full day planned. I plan to give a brief overview of the STI architecture, and then I'm going to turn the program over to our speakers, each of whom is an expert in his speaking area (Viewgraph 1).

Our STI Architectural Framework Working Group was established by our program director back in 1993 to address the questions that are involved in development of an STI architecture (Viewgraph 2). Identifying the current and planned STI program functions was one of the first things we had to address. We had to know what the program was doing in terms of developing an architecture, and even though identifying our functions may seem like a simple concept, it took a great deal of time and effort to do this effectively. We also addressed what components make up the current and planned STI data processing architecture, i.e., what we are using now to disseminate our information, to announce our information, to acquire our information, to exchange our information, and what we will be doing in the future.

Of course, this is tied in with our modernization plan, which was fortunately effective in gaining funding for modernization for the program. Another issue that we addressed was what standards exist that can facilitate the interoperation and interchange of the current and planned components. When I say components, I mean data, I mean information, as well as hardware and software, etc. Standards, of course, are critical in terms of interchange and interoperability, and in order to be effective for the STI Program or anyone else, they have to have a cost benefit associated with them (Viewgraph 3). If we standardize, we can cut costs and save money. We can also improve interoperability, scalability, and as I mentioned first, reduce life cycle cost. We can also simplify the management process.

The program has an Engineering Review Board that essentially overlooks all of the projects that are involved in modernization, including all of the procurements. The Board does periodic evaluations of what's being done in light of other projects, and essentially blesses or doesn't bless the plan that's under way. In order for the Engineering Review Board to do this effectively, an architecture is needed to guide the process. Although we receive funding for procurements, we didn't complete an overall architecture by the time we completed the modernization plan. This shortfall occurred because we had a time frame that was very short; the modernization plan was geared toward acquiring funding for modernization, and so we

wanted to get that plan out and get it done, and in fact, it was effective. As I said, we did get funding, but now we're going back to make sure that everything is going to fit together in the future and that we have an overall architectural plan. We also wanted to emphasize that this architecture is an important addition to our modernization effort and will be very important during our transition from the current systems to the planned systems.

Now, I'm not going to read the next slide to you, but if you want to take a moment and read it yourself, this is essentially a formal definition of the term "standard" that was put forth by the International Standards Organization (Viewgraph 4) (*A formal definition: "A technical specification or other document available to the public, drawn up with the cooperation and consensus or general approval of all parties affected by it, based upon the consolidated results of science, technology, and experience, aimed at the promotion of optimum community benefits and approved by a standardization body."*)

Of course, that's ISO and everyone knows that ISO deals with international standards. If you're done, I'll go on to the next slide (Viewgraph 5). Now this is very high level, and many of you in the room have been working on standards and so don't really need any definitions at all, but I thought I would mention that there are two major divisions among standards. One is the de facto standard, which is a specification of a product or system that has a dominant market share and which others tend to emulate. For instance, DOS is a de facto standard in the minds of some. There are other de facto standards that we have been looking at such as the emerging Adobe PDF format. Also, there are de jure standards, and these are standards that were brought forth by a standards developing organization, an accredited organization, such as ANSI, IEEE, AIM, and other organizations. Our government organization is NIST. We have within the de jure standards, for example, standards dealing specifically with hardware. For instance, we have standards that deal with electrical systems. Why are we able to plug in a lamp in any outlet in the United States and have it work? Because the plug conforms to an ANSI standard.

Now, going back to our architecture group, we essentially used a very standard methodology to work on the architecture (Viewgraph 6). First of all, we had regular meetings and in the minds of some of members, the meetings may have been a little too often, because it took quite a lot of time and effort to get their work items completed. As you'll note further down the list on my slide, we have work items completed by group members that were incorporated into a finished document by our technical advisor, which is MITRE, and every member of this "working" group essentially did work and produced a result that was presented to the entire group for approval. Essentially, all our decisions were group-oriented. They were by consensus. We had an emphasis within the group on the FIPS, because NASA, as a civil agency, must comply with the government standards put forth by NIST. We also emphasized industry standards such as IEEE's POSIX standards, NIST's APP, POSIX.O (OSE), and other standards such as the emerging standards that we think will be important to STI in the future (Viewgraph 7). In terms of the document that will come out of this group, we are planning a review of the document by NIST, but we need to get the proper approvals to have that happen. I mentioned the POSIX OSE reference model and noted that this was very important to us. In fact, the group started with this, and it was a way of looking at the essential



organization, of using common terminology such as applications software, application platform, external environment, interfaces, etc. I'm not going to go into anymore on POSIX OSE, because we are going to hear a speaker later today who is an expert in this area and will tell you everything that you want to know about it. So, at this point I want to turn the floor over to Dr. Lynwood Randolph, who is going to talk to you about OSI and TCP/IP. Dr. Randolph essentially is representing NASA in addressing questions related to this issue.

# **NASA Scientific and Technical Information Program**

## **– STI Architectural Framework Working Group –**

**April 13, 1994**



**Karen Kaye**  
**12th STI Program Coordinating**  
**Council Meeting**  
**kkaye@sti.nasa.gov**



## **STI Architectural Framework Working Group**

---

- Established by program director during summer 1993 to address questions related to development of an STI architecture
  - What are the current/planned STI Program functions?
  - What components make up the current/planned STI data processing architecture?
  - Do standards exist that can facilitate the interoperation and interchange of the current/planned components?

## **Value of Standards**

---

- **To be effective, must have an economic value ascribed to them**
  - **improve interoperability, portability, scalability, reduce life cycle costs, etc.**
- **Simplify the management process, particularly strategic planning**
- **Increased importance within STI Program during transition from present systems to modernized systems**

# **Standards**

---

## **A formal definition:**

**“A technical specification or other document available to the public, drawn up with the cooperation and consensus or general approval of all parties affected by it, based upon the consolidated results of science, technology, and experience, aimed at the promotion of optimum community benefits and approved by a standardization body.”**

## **Standards**

---

### **De Facto:**

**A specification of a product or system that has a dominant market share, and which others tend to emulate.**

### **De Jure:**

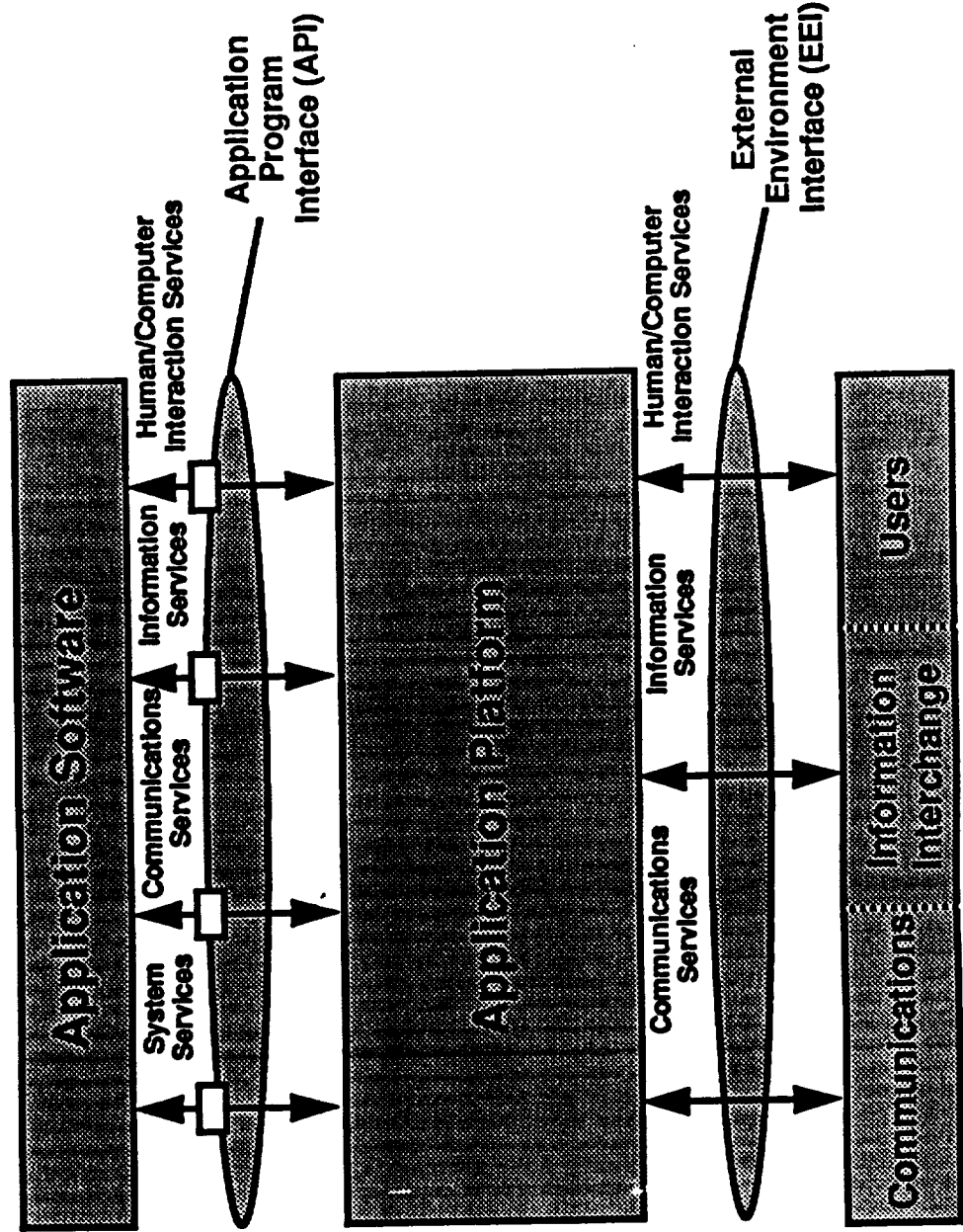
**The term applied to a specification created, approved, and maintained by an accredited standards developing organization.**

## **Group Methodology**

---

- **Regular meetings**
- **Decisions by consensus**
- **Emphasis on**
  - **NIST standards**
  - **Industry standards**
  - **Other**
- **Work items completed by group members and incorporated into finished document by technical advisor (Mitre)**
- **Planned review of document by NIST**

# POSIX OSE Reference Model





# OSI and TCP/IP

Lynwood P. Randolph, Ph.D.

First of all, I'd like to thank you for inviting me to share this information with you (Viewgraph 1). I've been on two sides of the issue with respect to Open Systems Interconnection Transmission Control Protocol/Internet Protocol (OSI TCP/IP). On the one hand, I've been on the policy side of NASA. I have helped to formulate and to pull together the agency's policy with respect to the open systems environment. On the other hand, I have worked very closely with organizations within the NASA environment that are required to implement this policy, and I have heard from them quite vigorously in terms of some of their concerns over the issue itself and how we can resolve it.

I'd like to first of all acknowledge the help and assistance of my colleague, Louise Goler-Brittain from Booz, Allen and Hamilton, who is here with us, who helped to construct the majority of this presentation, and I want to thank her publicly for doing that. Now, I'd like to start by, one, at least outlining to you how I will present the material that I have. First of all, the presentation is outlined in terms of some background information on the issue of OSI and TCP/IP. That's going to be followed by a statement of what the issue really is, and when I say really is, I mean the issue as it was, effectively, in September 1993, before a panel was appointed to publicly address the issue. I'm going to talk about something called the Federal Internetworking Requirements Panel (FIRP) report, its recommendations, NASA comments on those recommendations, what the next steps are, and finally a little about the draft report that I just received yesterday on the FIRP report.

Now a little about the background. The whole issue of OSI and TCP/IP is an issue that has been around for some time. About ten years ago, the Federal Government began to look specifically at open systems interconnection (Viewgraph 2). What would we have to do in order to put together a system that would be recognized world-wide for communications data, as well as information that has to do with video, as well as image? It was hoped at that time that OSI products would be produced in abundant numbers, that the vendors and such would continue to develop and support the development of products, and produce more products, and that the government and industry world-wide would make substantial investments in OSI. Also, there was a companion effort underway in the world of Internet. Incidentally, both of these efforts were funded by the Federal Government.

The Internet Protocol Suite (IPS) was a process that was associated with the Internet and it was a rather informal process in terms of standards. The development of standards generally requires a consensus which is documented by voting. But in the Internet process, this is rather informal and no specific voting is taken. The vendors were basically taking a wait-and-see attitude with respect to the Internet Protocol Suite. They weren't sure how much of the development was going to take place. And the primary drivers, at that time, were the Federal Government in terms of the defense and research industry, as well as some members of the research community itself. The TCP/IP suite itself had set up a standard protocol, and this IPS

runs on what's called a TCP/IP layer, which is comparable to the network layer for OSI. The report itself, principally a report focusing on the FIRP work, will concentrate on the Internet Protocol Suite and the network layer with respect to OSI. The Federal Government, in adopting and moving forward with the OSI standard, published what is known as GOSIP, which is the Government Open Systems Interconnection Profile (Viewgraph 3). It's the Federal Information Processing Standard 146. It was published in 1988, but before that time, there was a considerable amount of work performed to determine what OSI would become.

GOSIP, Version 2, was mandated by the Federal Government in October 1992, which meant that all procurements related to communications and applications that had functionality related to GOSIP, must conform to the GOSIP standard (Viewgraph 3). The GOSIP standard was expected to displace the IPS and other proprietary protocols. Incidentally, GOSIP, Version 3, is forthcoming. But the use of IPS continued to grow. It grew at a very rapid pace and, in fact, has reached the status where there are some 21,000 networks in the Internet, and there is reported to be something on the order of several million users. The development of IP products has proliferated. They saturated the market, much more so than OSI products. And to that extent, these products, the IPS products, are less costly and more readily available than the GOSIP products (Viewgraph 4). As a result, there is a need to develop perhaps more GOSIP products and to somehow give manufacturers the incentive to produce more GOSIP products. GOSIP products have not been widely implemented, as I have often stated, in Federal agencies. NASA is one of them; NASA invested in GOSIP products, but to a large extent have not used those; they've simply procured them because there was a procurement mandate, but they have not put them in use.

Instead, many of the Federal agencies, including NASA, have used exclusively IP, primarily because of their long history of operation. Well, this is somewhat of an embarrassment for the Federal Government, because the Federal Government is supporting both of these types of protocols. So, there was a formal need to address the issue (Viewgraph 5). In September of last year there was formally appointed an interagency group to look specifically at this issue and make some recommendations for the Federal Government to come forward. That's the essence of this report.

The group formally appointed to address this issue is known as the FIRP, the Federal Internetworking Requirements Panel (Viewgraph 6). It was appointed by NIST in September of 1993, and it had the strong endorsement of many of the interagency groups, including the Federal Networking Council and the Federal Information Resources Management Policy Counsel, FIRMPOC, which has its emphasis primary on GOSIP. FIRP members were to look specifically at requirements and make recommendations to NIST for resolving the conflict. The FIRP panel was established with nine members appointed. It is chaired by Diane Fountaine from the Department of Defense (Viewgraph 7). You will note that this is the panel that represents all Federal agencies. They were and currently are two NASA persons on the panel, Mr. Richard desJardins of the Goddard Space Flight Center and Mr. Milo Medin of the Ames Research Center. Both have been very active participants on the panel. You will note they are the only two from any single agency on the panel of nine members chosen by NIST to look specifically at this issue. I might point out that it's not coincidental that these

two gentlemen are on the list. They're world-wide experts in their fields, and they're recognized throughout the Federal Government for their contributions.

The FIRP took its duties very seriously and put forth a charter (Viewgraph 8). The charter was basically to look for both short-term and long-term internetworking issues. To make recommendations for the convergence of these two competing protocol suites, IPS and OSI, they looked specifically at the proliferation of proprietary protocols and some of the other related issues, which included the comparative strengths and weaknesses of both of these protocols. Neither one can actually sustain the operations as we would like to see them. What kind of support structure do we have for OSI as well as IPS? What's the role of proprietary protocol suites? What are some of the stringent security issues that are involved? What are the external relationships? We are not operating in a vacuum; NASA has several international partners who are primarily concerned with interoperation and communication aboard. So, what specifically would our international partners say as we make specific recommendations?

FIRP put forth a rather ambitious schedule of having its first meeting in October (Viewgraph 9). Moving right through the list, I see here on the chart, they are on track. They do have a final draft report that I just received yesterday, and they are on schedule with the recommendations coming forth. We will talk about those in just a moment (Viewgraph 10). The panel itself generally did not go off and do its own thing, but they interpreted their charter rather broadly and said, "Let's look at the whole process. Why are we doing this? What has driven us to this particular situation? Let's look at the process and let's move very promptly in terms of our charter and coming forth with some recommendations."

They outlined a report that they were going to produce (Viewgraph 11). They are going to talk specifically about requirements. What is it that drives these agencies? What's the principal driver for agencies in terms of communications? What about international interoperability for trade? How is that affected by these two protocols? Something about the standards process - how it operates, and how it will operate particularly with respect to two different protocols. Something about the technical issues; but they won't spend too much time on that. A little about economic considerations, and finally, the recommendations that they were going to bring forth.

One of the things they did was to determine the focus of Federal internetworking (Viewgraph 12). Basically, what are the requirements? They looked at the requirements primarily from the standpoint of process and structure. What are the core requirements necessary for agencies to communicate? They also identified something that had been reported earlier in a document by the National Performance Review and a very useful concept known as Affinity Groups. They looked at requirements, both from a functional point of view, as well as a characteristic point of view. Functionally, there are conventional kinds of things that must be considered - the requirements, messaging, information retrieval, transactions, composites. What are some of the characteristics of affordability, security, interoperability, accountability, manageability?

Finally, what specifically are these so-called "affinity groups." Affinity groups were named specifically in Vice President Gore's National Performance Review as basically groups which

have functional interests, share information electronically, and possess common information technology requirements. Thus, an affinity group is, for instance, the ICCN, which is the NASA Interagency Council on Computer Networks. The FNC is an affinity group, and you can think of a number of affinity groups. STI could be, and is, an affinity group sharing information electronically. There is a lot of emphasis on affinity groups within the report, and this is just a brief definition of what an affinity group is.

One of the broad interpretations of the requirements involves the vision for information transmittal (Viewgraph 13). The FIRP looked upon the information infrastructure as providing a seamless way of communicating throughout the country, both within Federal agencies, as well as among Federal agencies and the public at large. The FIRP believed very strongly that there has to be, and there must be, strong leadership if this is to take place, and not only in terms of words, but also in terms of deeds. The FIRP put strong emphasis on the OMB for that leadership because of its resources, policies and its oversight function. In terms of policy, OMB produces many regulations and guidelines that document how Federal agencies operate, and of course OMB does have significant oversight for information technology. The FIRP believes very strongly in its recommendations that the OMB should take a strong leadership role in this information technology interchange, primarily in resolving the issue with respect to TCP/IP and GOSIP, or OSI.

Next, a few words about the international situation (Viewgraph 14). The FIRP (and this is the report from the nine-member group) believes very strongly that international interoperability comes primarily in having Federal agencies that operate on the international level "do the right thing." They should work specifically in terms of their missions. They should work with their fellow foreign partners to select mission critical choices. Those choices that will enable them to carry out their missions very effectively. Agencies, as well as the affinity groups, should work closely together with their international counterparts in developing and fostering trade and communication. FIRP believes very strongly that successful international interconnection produces products and services that are available and are recognized internationally. They note the importance of basing Federal internetworking on internationally available technologies and solutions. The panel also strongly recommends that there be more participation by the Federal Government in industry consortia, and standards organizations. To a large extent, the Federal Government is taking a back seat in terms of these organizations, primarily because of the lack of resources for travel; but unless you are on the forefront, and unless you are involved in the standard organizations in producing and developing the standards, you simply aren't only going to react when those standards come forward.

One of the more controversial recommendations coming out of the FIRP report, involves the standards process itself and the so-called hierarchy of standards that are recommended by this particular group (Viewgraph 15). The first is not too controversial: they feel very strongly that in terms of hierarchical standards, the first standards should be open, international, voluntary standards developed by a development group. Second, the national voluntary standards or consortia standards, should be those that are considered. But the third, the proprietary or common use standards, should be considered. This recommendation has received considerable

criticism from a number of parties, both on the international, as well as the national front, and I'll talk a little about this near the end in terms of whether or not this particular recommendation will go forward. I also point out the international standards are rather formal.

The ISO, which is the International Standards Organization and the ITU, which is the International Telecommunications Union, sanctioned by the United Nations, are formal standards-developing organizations; they have formal membership; and have formal documented processes for developing standards. On the other hand, the IP Suite is supported by a group known as the Internet Engineering Task Force (IETF), and that task force also develops standards. But this standard development process is less formal than that recognized by international standards bodies. There is a great deal of concern over this informal mechanism that IETF uses in developing its standards. One recommendation that will be made is that the IETF's process be adopted and used for developing Federal standards. At the end of the report, I'll point out that this recommendation was not taken in the final report, and it will not be adopted in this particular format.

It is felt very strongly by the group that the GOSIP standard should be accepted (Viewgraph 16). It recognizes many of the benefits that GOSIP has. It also points out some of the shortcomings. A shortcoming is that this standard must be used in all situations and that other competing standards cannot be, or should not be recognized. The report takes exception to this; it supports the development and continued use of GOSIP, but it stresses that modification may be necessary in order for GOSIP to be more recognized and to be more useable in the future. The report also puts specific emphasis upon this. I see one of my colleagues here from NIST, who I'm sure is taking notes (Viewgraph 17). The report points out that it really stresses that NIST should identify and formulate preferred standards. A hierarchy of standards should be considered on the technical merits, but also take into account the marketplace and the costs. There should be coordination with the most effective affinity groups, and these affinity groups are defined with respect to their common interest and common sharing information. It also stresses that the agencies, like NASA, should be more active in working with standards development organizations, attend more meetings, participate in helping to develop standards. It hopes the Government will develop and merge into a single interoperable standards-based interconnecting or internet-working environment; at least, that's the objective.

Let me just mention a word about some of the technical aspects of the report (Viewgraph 18). One of the conclusions drawn by the experts was that there is no single protocol that will satisfy all of the requirements that are necessary, neither the OSI protocol nor the IPS. And therefore, each Federal agency must look toward accomplishing its mission first, what it needs to accomplish its mission, the use of appropriate protocols to accomplish the mission, and the available resources - the available products, supporting infrastructure, and plans that exist within the Federal Government. You cannot simply mandate purchase of a given product if the product is not available or if the supporting structure is not there to provide the backup that you need. One recommendation with respect to the economics has to do with the fact that it is very difficult to assess the impact of one or the other of these particular protocols, primarily because none will satisfy all the requirements. So, there must be a mixture in terms

of putting together a system to satisfy the agencies' missions and needs. It is further pointed out that the future demand for OSI products is uncertain (Viewgraph 19). The demand for those products now is much less than that for IPS products, primarily because manufacturers have tended to bundle IPS products with their own products. So, you can buy a given computer with certain IPS products already on it, but you can't do that with OSI products (Viewgraph 20). There are, however, some OSI products that are being widely accepted in the industry, primarily the X.500 directory service as well as the X.400 mail service, so there are advantages to products of both protocols.

Let me skip over to recommendations. The FIRP comes forward with a series of six recommendations. The background for the recommendations is that there is a vision that the FIRP has for interconnection; it is to provide a full range of integrated communications, for voice, for data, for imaging, for faxing, for Federal agencies, both within Federal agencies as well as among Federal agencies and their trading partners. To obtain such a vision, it believes very strongly that there is integration across Federal agencies for internetworking (Viewgraph 21). It feels very strongly that there should be clearly-defined and formalized responsibilities for operational support for these particular evolving structures. And as such, the FIRP made five specific recommendations: the first simply states that the role of oversight and integration across the Federal agencies for internetworking should be strengthened specifically within the Office of Management and Budget, with the strong emphasis that the OMB should be the driver in this whole process. I will come back in just a moment to give you the final recommendations (Viewgraph 22).

These were recommendations that were in the draft report, which was sent out for comment; the comments have come back; and these recommendations have been revised. Before I finish, I will give you the final revised recommendations (Viewgraph 23). Number two, the role and responsibility for fostering these standards and assessing technology changes should be focused and strengthened through the Department of Commerce. Recommendation number three (Viewgraph 24). I won't read through that, it's a long one to read. You have a copy. You can read it just as well as I, but the emphasis here is that responsibility for infrastructure development should be the core responsibility. I believe very strongly that there is a tie-in in terms of making communications available through this particular vehicle. In recommendation number four, they put strong emphasis on GITS which is the Government Information Technology Services Working Group (Viewgraph 25). The fifth recommendation has to do with OMB Circular A-119, which has to do with Federal participation in developing voluntary standards (Viewgraph 26). It recommends that the policy, OMB A-119, should be revised to reflect a wider range of interests, specifically in the area of international standards for the purpose of internetworking.

Now these were the recommendations that were contained in the FIRP report. The FIRP report was completed and circulated for comment to Federal agencies, including NASA, and this was done in February of this year. NASA reviewed the report and the comments were principally provided by the ICCN, which is NASA's Intercenter Council on Computer Networking, the organization that has responsibility for networking. NASA's review of the report was that, overall, it was a very pleasing report (Viewgraph 27). NASA felt and

recognized some of the shortcomings that GOSIP had. It also made some recommended changes in terms of how GOSIP could be more useful and emphasized the primary role of Federal internetworking. That was NASA's overall comment.

We did have, however, some very specific comments (Viewgraph 28). Specifically, we felt that the report sometimes was more of a sales pitch for the IPS than it should have been, that perhaps there were strong components within the community for that. Also, we should look very carefully at this whole idea of recommending proprietary protocols. This goes against the grain of formal standards, and we were a little cautious about that. We wanted to point out that it should clarify specifically the roles of IETF and GITS with respect to interconnection and internetworking. And some of our specific comments were that we felt that the agency should be held accountable. We wanted to know how the FIRP proposes that this be done. The report mentioned the affinity groups (Viewgraph 29). We were really concerned with the exact role of the affinity group. What is an affinity group? How could an affinity group be used to formulate and develop the standards? NASA's comments were forwarded to NIST and made available for public review.

As it turns out, it was picked up by the news media and our comments were included in an article reported in the Government Computer News of the March 21 issue (Viewgraph 30). Reporters tend to put spin on a particular coverage. They excerpted from our report certain words and made it appear as if we were not wholeheartedly supporting the FIRP report. We are. But this is the way the article was written. It also pointed out the significant skepticism of NASA's comments. Well yes, we were skeptical, but we did support the report. Just a word of caution that whenever you put forth a report and whenever a reporter asks you specific questions, you must be at least knowledgeable about where that report is headed. In any event, we were reported in Government Computer News on our "lack of support" for the FIRP report.

Comments were received from a number of places, and in fact, eighty-one comments were received on this draft report (Viewgraph 31). Reportedly, six Federal agencies made comments, including NASA. The summary also states that the report was favored two-to-one by those making comments within the United States, but comments from outside the United States were three-to-one, opposed to the report, and their comments also voiced strong concern from the standards organizations, the ISO and ITU. Now those comments all went back to the panel. The panel reviewed those comments over the last month, and they have subsequently come forward with the draft copy of their final report. I received a copy of that just yesterday from one of the members of the panel. What the draft report simply states is that there were comments received from twenty-two private sector organizations and from twenty-eight individuals within the United States. The comments took exception with some of the report's emphasis on proprietary standards. It voiced particular concern over the proposition that the IETF should be considered equivalent to an international standards organization. This was something that the majority of those who responded did not subscribe to.

If you go back to recommendation number one, the final report will come forward with the recommendation which reads as follows: The role of oversight and guidance for integration across Federal agencies internetworking activities should be strengthened. After due consideration, the panel dropped the consideration for OMB. The role of oversight and guidance for integration across Federal agencies interconnecting activities should be strengthened. By whom? It simply should be strengthened. So, that's one recommendation. The next, recommendation number two, was not changed. Recommendation number three remained unchanged. Recommendation number four remained unchanged. Recommendation number five was changed completely. Recommendation number five has to do with the OMB Circular A-119. In effect, it has eliminated that particular recommendation and has come forward with the following statement: The current GOSIP policy should be broadened to include appropriate standards drawn from both the OSI and the IPS protocol suites. It simply states that there should be a combination.

Now the panel, in making those five recommendations and having reviewed the comments that have come from other organizations, came forward with a sixth recommendation that was not included in this original report, and that recommendation is as follows: The existing FIRP panel should review the final implementation plans for Federal internetworking that are developing as a result of these recommendations (Viewgraph 32). A steering group should be established to review annually the Federal agency's progress toward achieving the internetworking vision outlined in this report (Viewgraph 33). In essence, what the panel has said is that they feel that the recommendations that are coming out of this report should be implemented; there should be a plan put forward; and there should be an annual review of that plan. Basically, it simply asks for all Federal agencies to develop a plan for internetworking, for that plan to be coordinated, and for there to be an annual review of that particular plan.

Now, as a result of all of these deliberations, NASA has been very forthright in terms of what it proceeds to do. Before the final report was delivered, members of the ICCN proceeded to put together a working group to modify its existing plans with respect to GOSIP. NASA already has a management plan for GOSIP. It has completed implementation plans, but those plans just focus on GOSIP. The ICCN is now looking forward to incorporating the recommendations in the FIRP report and to make those recommendations known within the agency.

Now one question is, what's going to happen to the FIRP report? (Viewgraph 34) The report goes from the committee to NIST, to the Department of Commerce, and there the final decisions will be made in terms of sustaining the recommendations or not. We're not sure what's going to happen. We, as an agency, are tracking this activity along with many others, and we have an extreme interest in what's going to happen, but we think that we have at least made an effort to structure this activity by having our people involved reactively up front in the area of standards.



---

# OSI and TCP/IP



**Lynwood P. Randolph, Ph.D., Manager  
NASA Information Technology Standards Program**

**NASA Scientific and Technical Information  
Coordinating Council Meeting**

**April 13, 1994**

---

**BACKGROUND**

- About 10 years ago, Federal agencies began planning to adopt the Open Systems Interconnection (OSI) standards as their internetworking basis
- During the mid-1980s, the Internet Protocol Suite (IPS) standardization process was just beginning
- The IPS is a subset of the Transmission Control Protocol/Internet Protocol (TCP/IP)

---

## **BACKGROUND**

- The Federal Government adopted Federal Information Processing Standard (FIPS) 146, Government Open Systems Interconnection Profile (GOSIP), in 1988
- IPS standards continued to grow

---

## **CURRENT SCENARIO/ISSUES**

- **There is a need for more GOSIP-compliant products**
- **As a result, GOSIP standards are not widely implemented**
- **Some Federal agencies have invested in GOSIP-compliant products that they have never installed nor used**
- **Instead, agencies are adopting IPS technologies or implementing proprietary systems**

---

## **ACTION**

- A need was identified to address the issues
- An interagency group was formed to tackle the problem

**FIRP**

- The Federal Internetworking Requirements Panel (FIRP) was appointed by the National Institute of Standards and Technology (NIST)
  - September 1993
- FIRP was endorsed by
  - Federal Networking Council (FNC)
  - Federal Information Resources Management Policy Council (FIRMPoC)
- FIRP members represent Federal agency mission interests and requirements

---

**Federal Internetworking Requirements Panel**

---

**Diane Fountaine, Department of Defense - Chair**

**Jason Canon, Department of Treasury**

**Michael Corrigan, General Services Administration**

**Walter Houser, Department of Veterans Affairs**

**William Hughes, Department of Commerce**

★ **Richard desJardins, National Aeronautics and Space Administration**

★ **Milo Medin, National Aeronautics and Space Administration**

**Thomas Rowlett, Department of Energy**

**Stephen Wolff, National Science Foundation**

**FIRP**

***FIRP is chartered to address the following and make recommendations on:***

- Short- and long-term internetworking issues
- Convergence of the IPS and OSI protocol suites
- Proprietary protocols, where appropriate
- Related Issues



## **FIRP/SCHEDULE**

• Chartered by NIST	Sept 1993
• First Meeting	Oct 1993
• Held 8 Meetings	Oct-Dec 1993
• Draft Report for Comment	Jan 1994
• Comments Received	Feb 1994
• Final Report	Apr 1994

---

## **FIRP/PANEL APPROACH**

- *Interpret charter broadly, not narrowly*
- *Pick the process, not the winners*
  - *Move promptly*

---

## **FIRP/REPORT**

### **FIRP Report Outline**

- **Requirements**
- **International Interoperability and Trade**
- **Standards Process**
- **Technical Issues**
- **Economic Considerations**
- **Recommendations**

---

**FIRP/REQUIREMENTS**

---

- **Primary focus of Federal internetworking should be on meeting agency mission requirements**
- **Federal internetworking processes and structures must identify core requirements and enable evolution**
- **Affinity groups should be identified and empowered to build interoperability**
  - **Affinity groups are functional interest groups that share information, electronically, and have common information technology requirements**
  - **The NASA Internetworking Council on Computer Networking (ICCN) is an affinity group**
  - **The Federal Networking Council (FNC) is an affinity group**

**FIRP/REQUIREMENTS**

- **The FIRP vision is that the Government Information Infrastructure (GII) to support internetworking evolves as a seamless part of the National Information Infrastructure (NII)**
- **FIRP believes there must be increased leadership and integration across Federal internetworking activities**
- **This leadership should come from OMB, supported by NIST**

---

**FIRP/INTERNATIONAL INTEROPERABILITY AND TRADE**

---

- **Best approach to achieve international interoperability is to rely on agencies to “do the right thing”**
- **Successful international internetworking products and services should be recognized by their viability in the international marketplace**
- **Government should participate in industry consortia and standards organizations**

---

**FIRP/STANDARDS PROCESS**

---

- **Federal internetworking should be based on a hierarchy of standards**
  - First, open international voluntary standards
  - Second, national voluntary or consortia standards
  - Third, proprietary or common-use “standards” with a preference for those that enjoy multinational commercial prevalence
- **Federal Government should recognize Internet Engineering Task Force (IETF) standards as open international voluntary standards**

---

**FIRP/STANDARDS PROCESS**

---

- **Modify FIPS 146-1, GOSIP to**
  - **Recognize both OSI and IPS protocols as acceptable internetworking standards**
  - **Recognize specific consortia and proprietary standards in exceptional cases (if clearly in the Government's interest and agreed to by affinity group)**
  - **Modify the name of GOSIP to reflect a wider range of Internet protocols**
    - » **Government Open Systems Profile (GOSP)**
    - » **Government Profile for InterNetworking (GPIN)**
  - **Change applicability statement to reflect agency preference for profiles defined in revised policy**



---

## **FIRP/STANDARDS PROCESS**

- **NIST should identify Federal, preferred standards**
- **Agencies should participate in standards development organizations**
- **Ultimate aim is to converge Government into a single interconnected, interoperable standards-based internetworking environment**

---

**FIRP/TECHNICAL ISSUES**

---

- **No single protocol suite currently meets the full range of Government requirements for data internetworking**
- **Each Federal community will pursue their own mission requirements solutions without a specific technical solution being imposed on them**
- **Implementation guidance must be provided on
  - Available resources
  - Market assessment of products
  - Government infrastructure and plans**

---

**FIRP/ECONOMIC CONSIDERATIONS**

- **FIRP expects Panel users demands for IPS products to escalate over next several years**
- **Future demand for OSI products is unknown, FIRP expects the demand for OSI products will grow at a substantially lower rate than for IPS products**

**FIRP/ECONOMIC CONSIDERATIONS**

- **FIRP believes that IPS products, on the whole, are generally cheaper than OSI counterparts**
- **A key issue in life-cycle costs is to minimize the number of open and proprietary protocols supported**

---

**FIRP/RECOMMENDATIONS**

---

- **FIRP's vision is that Federal internetworking should be a seamless component of the NII**
- **To attain this vision we must:**
  - **Increase integration across Federal agency internetworking activities**
  - **Refocus policies and technology assessments toward a more integrated and rapidly evolving telecommunications infrastructure**
  - **Clearly define and formalize responsibility for operational support of evolving infrastructure**
- **FIRP made the following five strategic recommendations**

---

**RECOMMENDATION 1**

***The role of oversight and integration across Federal agency  
internetworking activities should be strengthened within the  
Office of Management and Budget.***

---

**RECOMMENDATION 2**

***The roles and responsibilities for fostering standards and assessing technological change should be refocused and strengthened by the Department of Commerce.***

---

**RECOMMENDATION 3**

***The roles and responsibilities for infrastructure development and operations to support all internetworking services from advanced research and development to leading edge to core/commodity services should be clearly defined and formally assigned through the Information Infrastructure Task Force (IITF).***



---

**RECOMMENDATION 4**

*The roles and responsibilities of affinity groups should be defined, including how they are created and coordinated, by the Government Information Technology Services (GITS) working group.*

**RECOMMENDATION 5**

***In accordance with OMB Circular A-119, Revised, October 1993, voluntary standards should be adopted and used by Federal agencies, and international standards should be considered in the interests of promoting trade. The current GOSIP policy should be modified by the Department of Commerce to reflect the wider range of international voluntary standards for internetworking.***

**NASA COMMENTS/OVERALL COMMENT**

**NASA submitted comments to NIST in February 1994**

***Overall, NASA is pleased with the report.***

**NASA:**

- Recognizes the need for a GOSIP program without the mandate of a specific protocol
- Recommends changing the direction of GOSIP to achieve network interoperability
- Emphasizes that primary focus of Federal internetworking must be on meeting agency mission requirements

**NASA COMMENTS/GENERAL COMMENTS**

- The report sometimes seems to be a sales document for IPS
- The report should be more cautionary about accepting the possible use of proprietary protocols
- The report should provide clarification about the IITF and GITS organizations due to their critical roles

---

**NASA COMMENTS/SPECIFIC COMMENTS**

---

- **Executive Summary/Conclusions states “...agencies should be held accountable for meeting their mission objectives in as compatible a way as practicable with that vision.”**
  - **How will this be accomplished?**
- **Affinity groups are discussed in general but will be assigned the lion’s share of technical direction responsibility for Federal internetworking - from which policy will be developed**
  - **What is the exact role of affinity groups?**

---

**NASA COMMENTS**

---

***NASA's comments reported in Government Computer News***

- GCN, "Agencies Question Wisdom of Opening Up GOSIP", 3/21/94
- None of the Federal agencies that submitted comments on the FIRP report endorsed the findings, wholeheartedly
- Lynwood P. Randolph, IT Standards Program Manager, and Darwin Brown, ICCN Chair, were recognized in the article
- Significant skepticism was noted including NASA's commendation of FIRP's work but criticism of singular focus on IPS as solution
- NASA was quoted "[GOSIP should be a] structure of policies, guidelines, and procedures that encourage interoperability..."

**NEXT STEPS**

- **Comments received from public review have been analyzed**
  - **81 Comments Received**
  - **Only a Few Federal Agencies Commented**
  - **U.S. Organizations Were 2 to 1 In Favor**
  - **Comments From Outside U.S. Were 3 to 1 Against**
  - **Strong Concerns From Standards Organizations**

**NEXT STEPS**

- **FIRP Panel is undertaking the following reconsiderations**
  - **Coexistence, interoperability, convergence**
  - **IETF recognition**
  - **Security**
  - **Mandatory vs. preferred**
  - **Testing**



**NEXT STEPS**

- **FIRP Panel is undertaking the following reconsiderations**
  - **OMB, Commerce, IITF, GITS roles**
  - **International trade**
  - **NII**
  - **Steering role of FIRP Panel**

**NEXT STEPS**

- Final report to NIST is expected in April, 1994
- Proposed implementation actions, within the scope of the Dept. of Commerce and/or NIST, will be published in the Federal Register
- Issues related to Federal policy or regulation will be referred to OMB for consideration and action

# A Concise Introduction to MARC

Randall K. Barry

MARC, the acronym for MACHine-Readable Cataloging, is a term that has come to mean different things to different people in relation to library automation (Viewgraphs 1 and 2). Although it traces its origins back to a pilot project involving a small number of libraries, it is now almost impossible to touch upon automation in libraries without somehow involving MARC. Its use has expanded beyond libraries to a growing number of related institutions and professions.

Despite MARC's expanding use, not all professionals dealing with it understand what it is or why it's important. Many people who think they know what MARC is do not know that MARC is not a system; it is not cataloging rules—it is a data record structure. In order for library managers and automation specialists to make wise decisions on the choice between different MARC formats and MARC-based systems, they must become "MARC literate". The explanations that follow provide a concise introduction to MARC. They cover the elements of the MARC record, the formats that have developed around them, MARC's function in various institutions, and related topics such as its relationship to other standards, for example, Standard Generalized Markup Language (SGML). I've geared my treatment of MARC to those who may be unfamiliar with it and perhaps even with data processing.

My presentation will not provide all the information needed to actually work with MARC records or systems. To do that requires study of the MARC formats themselves and hands-on training with a MARC-based system. What I hope to provide is the groundwork for understanding MARC and a bridge to the technical MARC documentation that I'll mention at the end.

A MARC record consists of three basic elements: the record structure, the content designation, and the data content of the record. The first of these, the record structure, refers to the standardized way the information is organized. It follows agreed-upon principles and a finite set of encoding rules. The MARC record structure was originally developed as part of a library automation effort funded by the Council on Library Resources in the mid-1960's. The *MARC Pilot Project*, as it was called, was led by the Library of Congress and involved 16 other libraries of various sizes and types that wanted to encode their catalog information in machine-readable form and exchange it with others. The data structure developed for use in the MARC Pilot Project went on to become an American national standard (ANSI Z39.2) in 1971 and an international standard (ISO 2709) two years later.

The primary design characteristic of MARC is the division of character data into variable-length records (Viewgraph 3). Many other (non-MARC) data structures are designed around a fixed-length record (Viewgraph 4). Since the amount of information recorded from one bibliographic item to the next typically varies, a fixed-length record structure did not suffice. Internally, MARC records are composed of both fixed and variable-length fields,

however. The record structure supports definition of fixed-length elements below the record level for those pieces of bibliographic data that lend themselves to fixed-length data elements. (A majority of fields defined in MARC are variable length.) Fields may be subdivided as well into one or more fixed- or variable-length subfields. The MARC fields and subfields contain the actual data content that is gathered according to other standards (such as cataloging rules).

As I already mentioned, a MARC record consists of three basic elements: the record structure, the content designation, and the data content of the record (Viewgraph 5). I'm going to spend a few minutes talking about each of these in more detail.

A characteristic that distinguishes the MARC record structure from other data structures is the way fields are presented and referenced in the record (Viewgraph 6). Each MARC record begins with a special fixed-length field called the "Leader." Following the Leader is the Directory listing the names (tags) of other fields in the record. The last portion of the record is the variable field data. This segregation of the MARC record data into three structural components makes it easier to update a record when fields and subfields are added, modified, or deleted. The three parts of a MARC record are defined as follows:

**The Leader:** a fixed-length field consisting of 24 character positions, occurring at the beginning of the record. It contains important record-level information identified by relative character position and has no tag, indicators or subfield codes.

**The Directory:** contains the three-character tag, four-character field length, and five-character field starting position (relative to the first character position following the Directory itself) for each field in the record. The Directory follows the Leader with no preceding separator. Each entry in the Directory is made up of 12 characters; thus, the length of the Directory, although not fixed, should always be a multiple of 12. A special control character (hex 1E) signals the end of the Directory and the starting point "0" from which field locations are calculated.

**Variable fields:** the other data in the record are encoded in variable fields in the area following the Directory. (Note: It is possible to define fixed-length fields for this area as well.) This usually constitutes the largest part of the record in terms of number of characters. Each field in this portion of a MARC record ends with the same control character as the Directory (hex 1E); thus, this character is generally referred to as the "end-of-field" character. The end of the entire record is signaled by the "end-of-record" character (hex 1D). (Note that the end-of-record character does not replace the end-of-field character at the end of the last field.) Viewgraph 7 illustrates the same bibliographic information shown in some of the earlier slides. This time, however, the data are formulated according to the true MARC record structure.

To allow fields (and subfields) in MARC records to vary in length, they are marked explicitly by what is referred to collectively as the "content designation." Fields are identified by three-character tags. Tags are usually numeric, although the MARC structure does not limit

tags to numerals. (No implementation of MARC is known to have used a combination of alphabetic and numeric characters in a single tag, however.) (Viewgraph 8) Tags may be further qualified by alphanumeric characters called "indicators." Most implementations of MARC define up to two indicator positions associated with each tag defined. Indicators are placed before any other data in variable fields. Even when one or both of the indicator positions is undefined, blanks are usually supplied to reserve the space they typically occupy after the tag. This simplifies field processing.

Subfields in MARC records are identified by subfield codes, usually consisting of a single alphanumeric character following a special control character (hex 1F) called the "delimiter." Some of these control characters are used in other data structures as separators. MARC content designation is what most people think of when they think of MARC records, but content designation alone does not mean data conform to one of the MARC formats.

The MARC record structure and content designation defined to be used with it are the vehicles for transporting (communicating) data. Because the MARC record structure is so flexible, it can be used for all sorts of data. The data content of MARC records is character data of many kinds. It may be letters of the Latin alphabet, numerals, signs, symbols, special characters, letters of alphabets other than Latin, etc. The machine-readable data content of MARC records generally reflects the nature of manual information from which it is encoded. The data content of a MARC record captures the same information as the source, often enhancing it by the explicitness of the associated content designation (Viewgraph 9). (Information available implicitly in printed documents, e.g., indentions indicating paragraphs, are marked explicitly in MARC records.)

MARC content designation makes it possible to eliminate some characteristics of printed information when it becomes the data content of a record. Bold-face type is not carried over in the data content itself, but can be associated with the related content designators used. The most important function of the data content of a record is the by-products it supports. The words, titles, phrases, sentences, names, codes, etc. in MARC records are used by MARC-based systems to provide access (retrieval) and produce output products (display and print products). The structure (Viewgraph 10) and content designation (Viewgraph 11) alone would be of little use. The data content of a record would be equally useless without the structure and content designation inherent in MARC records (Viewgraph 12). All three elements are essential to the usefulness of MARC records.

A MARC format defines the list of valid data elements for specific types of records (Viewgraph 13). One format may contain specifications for more than one related record type. In USMARC, there are currently five formats defining 20 different record types. The list of data elements is usually organized according to their occurrence in the MARC record; thus, data elements found in the Leader are described first, followed by the Directory, and finally the variable fields. Since so many implementations of MARC use numeric tags, variable fields are generally organized numerically in ascending order, although they may not be used in that order in MARC records.

Variable fields are usually grouped into large blocks based on the highest order digit. In most of the existing MARC formats, the highest order digit broadly categorizes the kind of data that can be expected in the field (Viewgraph 14). The definition of format blocks in the MARC format for authority data is consistent with the bibliographic format for many field groups. The definition of blocks in the other USMARC formats shows some of the same consistency to blocks in other formats. In principle, new USMARC formats attempt to use the same field tags for defining data elements similar to those in other formats.

Below the field level, some parallelism in the definition of specific tags is also seen in most MARC formats (Viewgraph 15). Specific digits may function as mnemonic devices, regardless of the higher-level digit in the tag. Strict application of this principle is difficult, however, because it forces many available field tags to be reserved. The best example of parallelism in tag definition is in the 1XX, 4XX, 6XX, 7XX, and 8XX fields for bibliographic records and the 1XX, 4XX, and 5XX fields for authority records. In these fields the second and third digits have the same meaning in fields with different highest order digits. For other field groups, this kind of parallelism was abandoned in favor of using the fields for other kinds of data. For local information, MARC accommodates the definition of local data elements by reserving the digit "9" in tags and subfields. Most MARC systems make use of locally defined data elements (Viewgraph 16).

A growing number of record types have been defined in numerous MARC formats to accommodate a variety of data types. The number of MARC data elements in these formats is varied and large. Various types of material can be accommodated. (Viewgraph 17). Despite this, the number of data elements that are heavily used is rather small. People who use the MARC formats regularly become very familiar with the subset of data elements they need all the time. Most MARC records include only a fraction of the total number of available data elements. When the occasional need to encode unusual information arises, only then are MARC users usually forced to refer to MARC documentation. Library catalogers in particular find themselves "speaking" in terms of MARC tags after very little exposure to them.

The first use of MARC was for bibliographic data in the United States and Canada (Viewgraph 18). Five separate MARC formats for different types of bibliographic records (books, serials, maps, music, and films) were developed before being finally combined into one consolidated format with new content designators for two other types of material (computer files and archival materials) added in the process. By the time of that consolidation into a one-format document, use of MARC had spread to every continent except Antarctica. In most cases, the U.S. version of MARC was used as a model, but some liberties were taken with the data elements defined. Thus, it was necessary to begin referring to the MARC used in the U.S. as USMARC, to differentiate it from other "dialects" of MARC. The MARC record structure is common to them all, but the formats (i.e., lists of valid data elements) can differ.

Even after the consolidation of the seven USMARC bibliographic format specifications into one document, USMARC retained some of the separateness of the early years until format integration (approved in 1988 and being implemented) made any data element valid for any

type of bibliographic item. Foreign (i.e., non-U.S.) implementations of MARC still vary from fully integrated formats like UNIMARC to separate formats, like the MARC formats used in Russia.

After the first few years of use of MARC for bibliographic data, the same record structure was used to develop formats for authority data. Many of the bibliographic content designators (i.e., tags, indicators, and subfield codes) were applied to new types of data. Since then three additional USMARC formats have been developed in the U.S., accounting for the MARC format for holdings data, classification data, and community information. (Viewgraph 19). Recently, experimentation has even been done with using the MARC record structure for full text. I'll mention the experience of trying to apply the MARC record structure to full text in a few minutes when I talk about the relationship between MARC and SGML.

It is important to note at this point that MARC should not be confused with cataloging rules such as the *Anglo-American Cataloguing Rules* (2nd edition). Cataloging rules or other guidelines are applied in the formulation of information, whether in print or machine-readable form. Although elements of cataloging rules and information gathering policy do effect the way MARC formats are used, a conscious attempt has been made to keep the design and maintenance of the MARC standards detached from cataloging rules and policy.

Hundreds of vendors now market MARC-based computer systems. Systems are available that run on most platforms, from micro (PC) systems to large mainframe computers. As the variety of MARC record types and formats has increased, many vendors have enhanced their systems to accommodate wider uses. In fact, the desire of some vendors to expand their markets has helped to push MARC into new industries. These systems can be used to create and/or process MARC records.

MARC is pervasive in libraries. Most large libraries take a large portion of their cataloging from MARC record suppliers called "bibliographic utilities." The most well-known of these are OCLC (the Online Computer Library Center), RLG (the Research Libraries Group), and WLN (the Western Library Network). In Canada, ISM now serves as a major bibliographic utility. OCLC and several European networks are now vying for dominance in Europe. The situation in other parts of the world is not as clear.

Current MARC-based systems provide users with a lot of functionality not present in earlier systems. The first MARC-based systems functioned primarily to produce printed library cards and communicate bibliographic data to other libraries. Now MARC provides compatibility between different information systems, sometimes allowing organizations on opposite coasts to search each others' databases. At other times, MARC simply allows one organization to migrate its own data to a new system internally without having to massage or convert data to a new format.

Once institutions have their data in a MARC format and loaded to a MARC system, most find the design aspects of MARC improve retrieval and the ability to produce output products.

Many institutions undertake substantial improvement to their data at the same time as conversion to MARC.

The key to moving to MARC is the selection of a MARC-compatible system. The selection of a system should be done with a few basic functional requirements in mind (Viewgraph 20). The system should be able to import and export MARC records. It should be able to create, modify, and delete MARC records. The movement of data from one database to another, or one MARC-based system to another, should not result in any data loss. This is generally called "round-trip compatibility." Lastly, a MARC bibliographic system should be able to handle the bibliographic character set. There won't be time to talk about the USMARC character sets today, but I want to at least mention that there are unique features in bibliographic character sets which MARC-based systems need to be able to handle, particularly in the area of special characters and modified (accented) letters.

The challenge of conversion to MARC should also not be underestimated. Since the MARC formats involve a high level of data "granularity" (many pieces of information marked explicitly), some databases do not lend themselves as easily to conversion to MARC. Character set problems are also encountered, particularly with foreign language data involving accented letters. Fortunately, many MARC system vendors provide extensive conversion services, and some data lacking in a source file can be generated by default during conversion. Since there is a lot of competition in the market place, the cost for these conversion services has been going down. They also offer training for staff who may know nothing about MARC.

There is a common misconception that the encoding of data using the MARC formats will soon be replaced by Standard Generalized Markup Language (SGML), the highly successful standardized approach to encoding full text. This misconception is based on the observation that most documents encoded using SGML contain information that is bibliographic in nature. The SGML data elements (that is, tags, entities, etc.) used in the header and front matter of a full-text document often have a relationship to the MARC data elements defined for similar information in bibliographic records. Although there are similarities between SGML and MARC, those who jump to the conclusion that MARC can be abandoned in favor of SGML are overlooking important differences in the design and intended use of each standard.

SGML and MARC are alike in that they provide a standard structure for machine-readable information. This structure facilitates the maintenance and exchange of information. Each standard is non-proprietary, which means that they can be implemented without having to pay a royalty to the original developers. (Off-the-shelf SGML and MARC implementations are for sale, of course.) There is quite a market for SGML and MARC hardware and software. MARC (ISO 2709) and SGML (ISO 8879), as standards, facilitate the exchange of information between divergent systems and provide the basic framework for bibliographic and full-text data that have gained world-wide acceptance and use. Conformance to standards increases the marketability of products and promotes the exchange of information among a variety of sources.



SGML and MARC are different, however, in some of the functionality they were designed to support (Viewgraph 21). MARC was designed for large numbers of brief records. SGML, on the other hand, was designed to accommodate large quantities of data contained in single "instances" (documents). The structure and syntax associated with SGML-encoded documents was designed to make the processing of full-text data system-independent. Any SGML-smart system (that is, an application capable of interpreting an SGML Document Type Definition (DTD) and instances of its use (documents) conforming to a specific DTD) should be able to make sense of the structure and content of an SGML-encoded text. Depending on the level of markup, the SGML encoding can support a wide variety of print and/or display features. SGML markup will also support context-sensitive retrieval, based on indexing of data encoded with specific SGML tags.

SGML is highly hierarchical, with many tags occurring within other tags. MARC is less hierarchical, with little embedding of data elements inside one another, except that fields do contain subfields. The hierarchical nature of the full-text markup allows systems that are processing documents to make indexing decisions based on the relative importance of words and phrases that appear at various hierarchical levels. The ability to identify the hierarchy of text in a document is generally minimal in traditional word processing formats, thus the leap toward SGML. In terms of standardization of implementations, SGML is still young. Since SGML is only a structural standard, standard implementations of that structure are needed for various document types. At present there is considerable duplication of effort in the development of SGML DTD's and systems.

Although the MARC record structure was developed for different kinds of data than SGML, there is certainly overlap in some areas. As already mentioned, MARC was designed for cataloging data which are typically concise and dense, packing a great deal of intelligence into a small number of characters. The average MARC record is only 1,500 characters, functioning as a surrogate for the cataloged item. SGML was designed to full-text documents, which for even the shortest involves many times that number of characters (and perhaps image data). The MARC formats, which are implementations of the standard MARC structure (ISO 2709), define data elements designed to make optimum use of small amounts of data in a machine environment. These data elements easily support the functional requirements (print, display, retrieval) of bibliographic data. Proportionally more MARC data elements (tags, subfield codes, etc.) are designed to support indexing and retrieval of bibliographic data than are found in SGML data elements, where a majority of tags supports display and output requirements.

The precision and consistency needed for cataloging data has promoted the development of standardized cataloging rules for both description and choice of access points. These rules are reflected in the implementations of the MARC record structure which are also highly standardized. In the United States, only one implementation of MARC is used—USMARC. Other implementations exist (*e.g.*, UNIMARC), but they do not enjoy the world-wide acceptance of USMARC or the support of so many national libraries and computer system vendors. This high level of acceptance of one implementation of a data structure is one of the reasons MARC is so successful. Anyone with a MARC system can usually read in and

process USMARC data. The capability to import and export bibliographic data in some standard MARC format is almost always provided by systems. Full-text systems do not enjoy this level of standardization, and will not, even with the advent of SGML, until a small number of implementations of SGML have become well-established.

At some point it will certainly be possible to convert the structure of bibliographic data from a MARC encoding to SGML. Work has already been done to develop an SGML DTD (Document Type Definition) for the USMARC data elements. At present, however, MARC users have felt no pressing need to change the way bibliographic data are encoded or processed. Newly defined MARC data elements now provide links from MARC records to full-text SGML documents (or other non-bibliographic entities, like image data or audio). Libraries may never have to seriously consider any encoding for bibliographic data other than MARC.

So far this discussion of MARC and SGML has been from one point of view: data currently encoded following the MARC structure might be encoded using SGML. It's also worth mentioning that it was suggested that the opposite might also be feasible; that is, encoding full-text using the MARC structure. There was even a pilot project several years ago to use the MARC record structure to encode full text. A tentative MARC format for full text was designed, and portions of an important library text (*Anglo-American Cataloguing Rules*, 2nd edition) was even converted. Certain basic features of MARC did not make it well-suited for large amounts of text, however. It also became clear that certain limitations in the design of MARC made SGML far more suited for encoding full text, the inability in MARC to embed tags within other tags and the maximum MARC record length of 99,999 characters being the most noteworthy.

It appears that both MARC and SGML have their own niches in the computer age. The two have shown themselves to be compatible, although designed for different applications. It is important that experts in each structural standard be aware of the needs and uses of the other so that library materials in machine-readable format and bibliographic information about them can be easily integrated.

Machine-readable cataloging, and thus MARC, was the direct result of a crisis in libraries in the early 1960's. Libraries, particularly large ones like the Library of Congress, were having increasing difficulty keeping up with the distribution and filing of printed catalog cards, were experiencing reduced success in maintaining the alphabetical arrangement of the cards produced, and were running out of room for catalog card cases. MARC solved those problems and provided the means to vastly increase services to library users.

It is the standardization of and the conformance to the MARC record structure and formats that now allow data to be exchanged between systems and facilitate data use, storage, movement, and processing. MARC has also been a cornerstone in the development of networks and intersite information retrieval. Those that can provide their data in the MARC format open doors to potential users and facilitate participation in the growing global information community.

MARC has proven to be an invaluable vehicle of standardization, not only of bibliographic data, but of data in general, character sets, and networking. Adoption, or at least development of interfaces with MARC, prove ultimately valuable to any organization that has information to share. The Library of Congress' Network Development and MARC Standards Office is the maintenance agency for the USMARC formats (Viewgraph 22). The office is also the focal point for work on MARC in general and coordinates a USMARC advisory group called "MARBI" that meets in conjunction with the American Library Association twice a year. The Office also represents the U.S. internationally in IFLA and other organizations that work with standards for library information. It publishes the five USMARC formats, as well as six USMARC code lists, and other USMARC documentation. For more information on MARC and related documentation, don't hesitate to contact the office in Washington.

# **A CONCISE INTRODUCTION TO MARC**

Presentation by

**Randall K. Barry**  
**Senior MARC Standards Specialist**  
**(Internet: rbar@seq1.loc.gov)**

**Network Development and MARC Standards Office**  
**Library of Congress**  
**Washington, DC 20540-4020**

***MARC-1 Introduction***

**MARC = MAachine Readable Cataloging**

- Encoding of information into *machine-readable* format so that it can be processed by computer
- Originally designed for the *bibliographic* information found in library catalogs
- MARC is not a *system*
- MARC is a data record *structure*

**MARC-2 *Definition***

# VARIABLE-LENGTH RECORDS (MARC-like):

\$ = Unused portions of fields      Variable-length records

100\$aLe Carré, John\$d1931	245\$aThe night manager :\$ba novel
-----------------------------	-------------------------------------

100\$aMessora, Noemi	245\$aCassell's contemporary Italian :\$b a handbook
----------------------	--

of grammar, current usage, and word power	
---	--

100\$aFollett, Ken	245\$aA dangerous fortune
--------------------	---------------------------

100\$aClancy, Tom	245\$aWithout remorse
-------------------	-----------------------

100\$aWouk, Herman	245\$aThe hope :\$ba novel
--------------------	----------------------------

100\$aUpdike, John\$d1932	245\$aRabbit at rest
---------------------------	----------------------

MARC-3 *Variable-Length Records*

# FIXED-LENGTH RECORDS (non-MARC):

\ = Unused portions of fields      Fixed-length records (66 characters)

Le Carré, John\	1931	The night manager : a novel\	////////////////////////////////////
Messora, Noemi\	\\\\	Cassell's contemporary Italian : a handbook ...	
Follett, Ken\\	\\\\	A dangerous fortune\	////////////////////////////////////
Clancy, Tom\\	\\\\	Without remorse\	////////////////////////////////////
Wouk, Herman\\	\\\\	The hope : a novel\	////////////////////////////////////
Updike, John\\	1932	Rabbit at rest\	////////////////////////////////////

MARC-4 Fixed-Length Records

## **MARC RECORD ELEMENTS:**

**A MARC records consists of three elements:**

- 1) The record *structure***
- 2) The *content designation***
- 3) The *data content***



## **MARC RECORD STRUCTURE:**

A MARC record's structure follows national (Z39.2) and international (ISO 2709) standards. It consists of three (3) structural components:

- 1) **Leader** - The first 24 character positions
- 2) **Directory** - 13-character entries (1 per field)
- 3) **Variable fields** - Of any number or length

# MARC STRUCTURAL MODEL:

LEADER	DIRECTORY
00722nam\\2200217\ a\4500	001001300000 008004100034 040001300100
100002500168 245010900193 260004400302 300003500346	
VARIABLE FIELDS	
93018041# 940411s1992\\nyua\\b\\000\0\eng\\# \$aDLC\$cDLC#	
1 \$aMessora, N.\$q(Noemi)# 10\$aCassell's contemporary Italian	
: \$ba handbook of grammar, current usage, and word power /\$cNoemi	
Messora.# \$aNew York :\$bMacmillan Pub. Co.,\$cc1992.# \$axil, 516 p.	
: \$bill. ;\$c24 cm.#	

MARC-7 MARC Structural Model

## TAGGED RECORD DISPLAY:

```

LDR 00722nam\\2200217\a\4500
001 93018041
008 940411s1992\\nyua\\b\\000\0\eng\\
040 $aDLC$cDLC
100 1\ $aMessori, N.$q(Noemi)
245 10$aCassell's contemporary Italian :$ba handbook
    of grammar, current usage, and word power
    /$cNoemi Messori.
260 \\ $aNew York :$bMacmillan Pub. Co.,$cc1992.
300 \\ $axii, 516 p. :$bill. ;$c24 cm.
  
```

## CATALOG CARD DISPLAY:

**Messora, N. (Noemi)**

Cassell's contemporary Italian : a handbook of grammar, current usage, and word power / Noemi Messora. -- New York : Macmillan Pub. Co., c1992.

xii, 516 p. : ill. ; 24 cm.

93-18041 DLC

*MARC-9 Card Display*

# MARC STRUCTURE ALONE:

LEADER	DIRECTORY
////////////////////	000000000000 000000000000 000000000000
00000000000 000000000000 000000000000 000000000000 nnnnnnnnnnn	
VARIABLE FIELDS	
??????????# ???????????# ?????????????????????????????????# ???????????#	
??????????????????????# ???	
??	
??	
??????????# ???# ?????????????????	
??#	

## CONTENT DESIGNATION ALONE:

001			
008			
040	\\\$a	\$c	
100	1\\\$a	\$q	
245	10\$a	\$b	\$c
260	\\\$a	\$b	\$c
300	\\\$a	\$b	\$c

**MARC-11 *Content Designation Alone***

## DATA CONTENT ALONE:

00722nam\\2200217\\a\\450093018041940411s1992\\\\\\n  
ya\\\\\\b\\\\\\000\\0\\eng\\DLCDLCMessora, N.(Noemi)  
Cassell's contemporary Italian :a handbook  
of grammar, current usage, and word power  
Noemi Messora.New York :Macmillan Pub. Co.,c1992.  
xii, 516 p. :ill. ;24 cm.

**MARC-12 Data Content Alone**

## **THE MARC FORMATS:**

- **A format defines...** the list of valid data elements for specific record types
- **In fixed-length fields...** the character positions;
- **In variable fields...** tags, indicators, subfield codes;
- **Requirements...** mandatory data and optional data;
- **Repeatability...** what is repeatable or not repeatable

**MARC-13** *The MARC Formats*



## **BIBLIOGRAPHIC FIELD GROUPS:**

- **0XX - Control information, numbers, codes**
- **1XX - Main entry**
- **2XX - Titles, edition, imprint, etc.**
- **3XX - Physical description, etc.**
- **4XX - Series statements**
- **5XX - Notes**
- **6XX - Subject access fields**
- **7XX - Added entries, linking fields**
- **8XX - Series added entries, holdings, etc.**
- **9XX - Local implementation fields**

**MARC-14 Bibliographic Field Blocks**

## PARALLELISM BETWEEN FIELD GROUPS:

Within some field groups several field tag patterns repeat with the same general meaning.

- **X00** - *Personal names*
- **X10** - *Corporate names*
- **X11** - *Meeting names*
- **X30** - *Uniform titles*
- **X40** - *Bibliographic titles*
- **X50** - *Topical terms*
- **X51** - *Geographic names*

## LOCAL DATA ELEMENTS:

All USMARC data elements including the digit "9" are reserved for local definition and implementation.

- **9XX** - *Local field group*
- **X9X** - *Local fields within a group*
- **XX9** - *Local field of a particular category*
- **\$9** - *Local subfield*
- **9** - *Local indicator or fixed-length data element value*

(Field 490 is the one exception. It is a valid field.)

## **USMARC BIBLIOGRAPHIC FORMAT:**

This list of content designators covers seven categories of material.

**Language material (Books)**

**Computer files**

**Maps**

**Music**

**Serials**

**Visual materials**

**Mixed materials**

## **THE EARLY MARC ERA (1965-1980):**

- 1) *Books, a MARC Format* (1970)**
- 2) *Serials, a MARC Format* (1970)**
- 3) *Films, a MARC Format* (1970)**
- 4) *Maps, a MARC Format* (1973)**
- 5) *Music, a MARC Format* (1976)**

(Several editions of some of these formats were published separately.)

## **THE CURRENT USMARC FORMATS:**

- 1) *USMARC Format for Bibliographic Data*
- 2) *USMARC Format for Authority Data*
- 3) *USMARC Format for Holdings Data*
- 4) *USMARC Format for Classification Data*
- 5) *USMARC Format for Community Information*

(Several editions of some of these formats have been published.)

## MARC-BASED SYSTEMS:

The following are some of the key functional requirements for any MARC system:

- 1) Ability to *import MARC records*
- 2) Ability to *export MARC records*
- 3) Ability to *create/modify/delete records*
- 4) "*Round-trip*" *compatibility* (i.e., no data loss during import/export)
- 5) Ability to handle *bibliographic character set* (i.e., "combining" characters/diacritics)

## **RELATIONSHIP BETWEEN MARC AND SGML:**

- 1) MARC--for large numbers of brief records**
- 2) SGML--for large quantities of full text**
- 3) MARC--exchange/printed output/retrieval**
- 4) SGML--printed output/exchange/hypertext**
- 5) MARC--can link to SGML files**
- 6) SGML--can embed or link to MARC record**

Each standard is designed for specific applications but share some common uses

**MARC-21 MARC vs. SGML**



## **MARC STANDARDS SUPPORT AT LC:**

- **Maintenance agency for USMARC:**  
*Library of Congress, Network Development  
and MARC Standards Office*
- **Coordination of advisory group for MARC:**  
*Semi-annual MARBI meetings (at ALA)*
- **Documentation: 5 formats; 6 code lists;**  
**specifications for record structure, character  
sets, exchange media; MARC record services**

**MARC-22 MARC Standard Support**



# Open System Environments

Fritz Schulz

Good morning, ladies and gentlemen. I am the manager of the Distributed Systems Engineering Group at NIST. Today, I want to talk about Open System Environments (Viewgraph 1). That's a tag line, it's non-parseable. It's a token; you have to take it as a sort of a brand label, and it means certain things to a number of communities that are scattered around both the project community and the standards community (Viewgraph 2). I am going to focus my discussion for a particular customer. I am going to characterize that customer by giving you a scenario. You're a project manager. You've just been given the job and you realize there are problems out there somewhere, but it's a nice day and you've just gotten your promotion, and you just have a shiny new badge and a clean desk for the first time in a long time - and the last time in a long time. You've been given the responsibility for integrating a number of information systems and making sure they work, and you've just begun to realize that some of those systems are not in your chain of command, but it's clearly in your job description that these systems must work together.

For instance, you have a system that is a repository of information and must be made available to a customer set. Most of those customers aren't in your information systems organization. They may be within NASA or some other part of the Federal Government. Indeed, there may be taxpayers out there or businesses, aerospace industries, that need to exchange data with you. As part of your responsibilities, you need to establish good coordination, interchange of data and so forth.

But there's no program mechanism for establishing that at this point, for establishing the capability for doing that. You can't mandate standards on this customer set. As a matter of fact, more often, they are mandating to you what standards you'll use, and in many cases, they don't line up with the technology that is already in the procurement pipeline of the project pipeline for your organization. You don't have the skill set currently deployed able to handle that kind of technology. What do you do? Well, you start up a Distributed Systems Engineering (DSE) program. There's a good answer; that's not your total answer. But it's the part that talks about consensus specifications and how to handle what we call standards. I want to create a little bit more of a spectrum of specifications that you need to deal with in your program.

Heterogeneous distributed systems (Viewgraph 3). What do we mean by heterogeneous? It means a number of things, but it's mostly about acknowledging reality. Heterogeneous means multi-vendor. It means your procurements will always have more than one vendor who can satisfy your procurement need; thereby, you get fair procurements. Multi-vendor also means that no single vendor is going to be able to supply your information technology needs over the breadth of the whole information system or a distributed system that's in place. As I have mentioned before, many of the systems that you'll be interacting with are in completely

different organizations, many of which aren't in the government, and they have their own procurement practices and their own procurement drivers as well.

So, you will never be in a position to assume what brand name is on the other side of that wire or interacting with your applications software. Heterogeneous means multi-vendor. Heterogeneous also means multiple kinds of technology. In any of five or six key technology requirement categories, there will be multiple solutions, each of which will carry its own advantages and disadvantages. Your organization will see fit to deploy multiple solutions in different places throughout your distributed system. You'll have to choose different technology because your needs will be different in spot areas. In some places, you might use OSI; in some places you might use TCP/IP; in other places, you will use ISDN, and there are a number of other solutions that are coming to the floor. We'll get to those in a few moments. We can put the reference model up, and I can wave and point and wave my hands around a little bit and give you some of the alternatives.

So, heterogeneous means at least multi-vendor and multi-technology incorporated in your distributed systems. Now, let's go and talk for just a moment about some definitions. These slides give some definitions, and I am not going to go through them in detail (Viewgraph 4). What do you mean by open systems? Well, that depends on your objectives. This was a heavily, heavily negotiated definition. But it does a couple of things for programs. My focus here is on establishing the tools a program needs to get consensus in place and establish some strategic directions. For example, a system that's sufficient to.... What's sufficient? We beg the question here. Sufficient *open specifications*. You expected to see standards there instead of open specifications, didn't you? No, you won't see standards there; there are real reasons for using the term *open specifications*. For interfaces, services and supporting formats, you saw an excellent presentation just a few moments ago that focused very well on the data format area.

Data formats are the only open-ended area that we see throughout this whole engineering approach to enable properly engineered applications software to do the following things. There are three capabilities. It must be ported. It must interoperate with other applications software scattered across the distributing system, and it must interact with users in a style which facilitates user portability. That's when people sit down at a computer. We don't want to have to retrain everyone; we want to be able to make some assumptions about what's happening out there across the Internet or wherever people are touching our data. We need to have some assumptions and conventions in place so that we can write our software so that they know how to interact and use our data and so forth.

These are programmatic level capabilities. There are many, many other capabilities that don't need to be established at the programmatic level, but one of the key objectives of DSE/OSE is to identify those questions which must be addressed at the program level. We wanted to find the questions that could not be addressed in isolation and package those up and to make decisions in those areas at the program level and then protect those other decisions - protected for the people who are trying to get the rubber against the road and get some work done and not bog them down with decisions that are taken at too high a level, for no apparent reason other than the fact that a decision was possible, so someone went ahead and did it. These

questions and these capabilities are ones which must be addressed at the programmatic level — properly engineered, by the way.

I skipped over that fairly lightly. You can do the right thing at the program level, but if you don't follow good engineering practice, you will not achieve portability; you will not achieve interoperability; and you will not have a common method of interaction with the users despite the fact that you mandated standards and that they are in use, in fact. Throughout organizations, you can actually have those standards be used and fail to achieve portability, interoperability and user interaction. So, we need to pay attention to both the specifications and the engineering practice. Now, we talked about open specifications. We didn't duck that question. We said they were public specifications that are maintained, not written — maintained by an open public consensus process to accommodate new technologies over time. Don't build ourselves into dead ends with standards that are inconsistent with international standards: let's migrate toward international standards where they appear.

Open specifications are a very, very important programmatic tool (Viewgraph 5). There are legally and politically defensible methods for applying open specifications to both procurements and engineering organizations. And they are an important mechanism that lets you address the problem that there aren't enough standards. Despite that there are a whole lot of standards, there aren't enough standards to address any single problem. You must go, kind of on a scavenger hunt, go out there and find specifications that meet your needs and leverage consensus where you find it. It may not be recognized; it may not be total consensus. It may be confined to a particular area, and there can be a lot of very good reasons, legitimate reasons, that you can't find the standard for something, but use that consensus where you can get it. We have a rule of thumb that we use in the OSE program: "Since consensus is so expensive, get all the consensus you can get for free and then only pay for as much consensus as you can get away with." What that means is, if you look at the total cost for a single standard, pick any standard. I don't even care what it is. Multiply out all the staff hours and the travel cost and the salary of the people and so forth and so on; you will find a number that will horrify you. We don't encourage people to do this, because it may guarantee that management will not support a standards committee ever again, and we need participation of the standards committee. But they are very, very expensive.

Use that consensus where it exists, because if you don't use it, you will have to duplicate that cost on your program. Go ahead and multiply that out and put it against your program cost. You will have to duplicate that cost for a smaller consensus set. And there is another fact of life: The larger the consensus set, the longer it takes and the more money it takes to get there. If you are aiming for a worldwide consensus on every standard that you are going to need, it's going to take you a long time and it's going to cost you a lot of money. You need to tune the level of consensus that you will accept on specifications that don't require international consensus. So, open specifications are a key cost saving tool for programs.

I present two other definitions, portability and interoperability (Viewgraph 6). Definitions are extremely important in three different forums. I won't say we wasted a year, but we found ourselves three different times, often with many of the same people in the room, having to go back a year and start over again to get a clear definition of what these things were, and

everybody knew what it was at the beginning. Portability, interoperability. We know what those things mean. Everybody nods and we move on. Well, we came back a year later and we wrangled for a while, and then we settled on our definition and moved on. These two aren't the only words that that kind of thing can happen to. There are other words that each different standards community or each different project is going to have to settle on, but people don't believe this. It sounds crazy, but you can hold yourself up a year by not paying a little bit of attention up front and making sure you've got clear definitions to some of these things. By the way, many of these people had these definitions reversed. That's how little consensus, how little understanding, we had at the beginning of our conversations.

Now, if you don't know where you are going, any road will do. This is where we were going. And the way it worked out, the process that we arrived at may be somewhat useful to you. Although I am not a big fan of process standards, and I certainly don't encourage you to do the same thing any other group is doing, because process has to be tailored to an organization, let me give you the flow that we used to arrive at this set. These two were important objectives up front (Viewgraph 6). This activity was pursued in a program context. By the way, NASA was very involved in the early days; this is a six-year project. This ballot that's going on at the international level has taken six, almost seven years, to get to this point. There is a lot of policy that is involved. Whether you want to use that word or not, it has large implications for how standards committees and organizations pursue their business and coordinate with other people. So, that's part of the reason that it took so long. NASA has been involved in this for some time. Many of the centers were involved. I'd like to get involved with you folks and make sure that you are aware of some of the things that came in critical times.

So, programs addressing needs, making a clear connection to user requirements, and being driven by user requirements were key objectives here. The rest of these are sign posts along the way (Viewgraph 7). They were added when we came to a fork in the road and said, "We can meet these objectives by going this way or by going this way." And each one of these is a sign post that says, "Well, we'll go this way." And this is a sign pointing to the direction that we took. For instance, you can achieve these by picking a single vendor. I mean, it's a trivial case, okay, but there are other ways to do this as well. What we said was, "We're going to intercept the standards process, and we're going to be working with the standards process to accelerate it or tune it or do what can be done with the standards process to make sure the right things get done." Accommodation of new technology — you can build yourself into a dead end very, very easily. It's surprisingly and shockingly easy to build yourself into dead ends, and we've seen that on a number of occasions.

Application platform scalability and distributed systems scalability are two objectives. We combined the bullets because we wanted to keep them to a small number. But this means that you can put many different kinds of platforms on a distributed system, and if that distributed system is designed in such a way, you can put on special purpose processors, such as tech search engines, or realtime process control for industrial automation, or for other kinds of process control.

Some of you may be involved in or may have been involved in some process control type activities. Distributive systems scalability is an attribute of the system as a whole, not of any given platform or any given communications media (Viewgraph 8). But there are many aspects that require architecting in the global sense or just below that level that need to be involved as well. So, we are aiming at both of those objectives. These were sign posts along the way, and we think there's a lot of lessons learned and insight that filtered into our work as a result of that. And those are our objectives, and this is the approach we took. Anybody in here a systems engineer? Or have systems engineering in their deep, dark past? Okay. You folks. This is Systems Engineering 101. Say there is an interface and there is a black box on Side A and Side B. These three words cost us more trouble than anything else that we did all the way through, including application portability and interoperability. These were little speed bumps along the way compared to these things.

I'm going to give you the snap definitions of these, and then I think you will see how confusing things got for a period of time. Interface is just a boundary. It's a place. It's that infinite plane between two things of zero thickness. It's just a place. Something that penetrates that boundary is called a service. It's exchange between two things or two entities on either side of that boundary. Now, a requirement. Remember, I said we were going to be driven by user requirements? A requirement is a statement of need for a particular service, at a particular interface. That's what a requirement is, okay? Nothing else is a requirement. There are other kinds of requirements, okay? But we label these service requirements to distinguish them from many of the other types of requirements. We say we need this service at this interface. That's a unitary requirement.

Now, if you have need for the same service at a different interface, that gives you a different requirement, and you may do other things in that other area to satisfy that requirement. Requirements are used to drive the selection of standards, and they stand alone and stand separate from a selection of standards. I've got a list of standards over here. I've got another book over here that gives me my requirements. A lot of people say, "Well, what are my requirements? My requirements are for Posix and for X-Windows, TCP/IP." I say, "No, no. We need to identify those requirements separate from the specifications, because these things evolve." We've all seen that, probably at close range, in the recent past. And in the immediate future, too. Because it goes on and on and on.

So, the programmatic principle is that you get a clear statement of user requirements and let it drive the selection of standards and define them separately. There will be different people in different organizations following different principles - updating and modifying your standards base and your requirements base. So, we've got our objectives; we know where we are going; and we've got our approach. Now, what interfaces, what specific interfaces, and what specific services are involved in meeting those objectives? Not many.

I'm going to flash this up here (Viewgraph 9). I'm not going to talk to this slide because it lies a lot, okay? It says, oh, there are two interfaces and three things. Okay. Two interfaces, just two. That's a number anyone can handle, right? Well, it's not true at all. There are four interfaces. Four interfaces. One, two, three, four. (Speaker indicates the four interfaces on the viewgraph.) Notice that there is a thing on each side of those interfaces. And we've

characterized the types of specifications; we've characterized the services; we've characterized the test technology that needs to be in place to validate and verify and cite performance to the standards that are associated with service requirements. Let me step through them very briefly.

Let's start with the application platform. Application platform is that box you pick up out of the peanuts in the cardboard box and drop it on your desk. You haven't loaded any applications on it at all. It's everything that is wrapped up in there, though. It's systems software; systems software is in here. Operating system, drivers, schedulers - all those kind of good things are in there. There's hardware in here. One of our major principles is for us, at this level, at the program level, is this: "Don't open that box!" Resist all temptations to open that box. People have opened that box and we have seen careers go down in flames. We've seen all kinds of interesting things happen when people open those boxes at the program level. Someone does need to open those boxes. Leave it to the people who are trying to get the job done. Let them pick the technology that they need to have. There is a wide variety of technology that goes in this box.

Application software. Everybody knows what application software is. You walk out of CompUSA holding it in that little bag, and you pull it out, and it is on a little disk. That's applications software, and a major principle of the program is to maintain a clear distinction and a clean interface between those two things, because you're going to buy some of these, and you're to build some of these, and these things need to work together, and especially in situations where what you buy needs to interoperate with what you buy. The API specifications are going to be what you're going to mediate. There are a lot of uses for the specifications that lie on this interface. Now, I have just described the two things that are on each side. We tend to think of this as a source code listing for applications programs. The reason that doesn't work for everything is because you buy a lot of these and you never see the source code. But that's okay, I don't mind. We'll get back to that in just a couple of minutes. I've got a source code listing, and I want to be able to write that software so that it will ride on this platform and work in a distributive environment.

What specifications do I give to my programmers to be able to protect my investment in that software? The answer to that question is what specification winds up on the API? Those books on that programmer's shelf — that's a very easy way to characterize it so that everybody understands what that is. There are other ways to characterize it. And that set of specifications has other effects besides being used for your programmers. It also characterizes the services that you need from your application platform. Say I am going to buy a platform. And I know what population of software is going to run on that platform today, because I am replacing something that's over here right now. But I don't know a year from now or two years from now what software is going to be running on that platform. So, I want to characterize that platform in terms of what services I think I am going to need from a strategic point of view, and I am going to commit my organization to a long time down the road. So, APIs also have a lot of implications for selection on its platform. Now, the other three interfaces share an interesting property. They are all accessible and are physically characterized from outside the application platform. You can see them when you walk up to your platform, if this is a platform that runs on a distributed system.



Let's start over here. Human-computer interface. These are people, you know. They are that really challenging element of any distributed system that gets you into the most trouble at any time, at any point, and they are very difficult to characterize. They are not linear and they do all those kinds of interesting things. What does a human-computer interface look like? What do the specifications look like? They look like human factor specifications. They say, "Here's a picture of what happens on the screen. And if you double click on this little region of the window, this is what a window looks like; if you double click on this little region, something happens." You characterize what happens, and then you describe this in a style guide type form. It's also more than that. It's any way that information technology interacts with people. There is a human-computer interface on the telephone; it's the buttons. And it's two-way audio. That will very carefully describe bandwidth that is available at that interface. It's audio, video; its mice, keyboards, display screens, everything - every way that people interact with the information technology is on this interface.

Let's go over here. These are world-class bad names and also a very good example of what happens when you have to negotiate something at the international level. These things are actually physical media. We're starting to get into the territory that our previous presenter covered quite well. These are physical media. They are physical things that you could pull out of any machine. It's your CD ROM; it's a cassette tape; it's a disc cartridge; it's anything that you could pull out and carry away the stored data. There is no protocol here — just data storage and retrieval. The specifications that sit on this interface are of two kinds. They form kind of a matrix. The first kind is physical media. And physicists get involved in this kind of thing — pencil, string, thickness of the oxide, the amount of magnetic flux it takes to flip a bit from Point A to Point B, etc., etc. All that kind of physical characterization — the size of the disc, and where the holes are in the center, and so forth and so on. Physical media. (Viewgraphs 10-20 are included for informational purposes.)

The other kind of specifications are media independent data format specifications. Media independent data format says, "This is what a document looks like," or "This is what an ASCII file looks like," or "This is what an audio string looks like," — or a video sequence, or a still picture, or a bibliographical record, and so forth and so on. There is a whole world of discussion that needs to go on just on that side of the axis on this interface. Those are two kinds of specifications: media specifications and media independent specifications. Media independent data format specifications.

Come on over here. Communications services. This is protocol. And everything that is out here is all the infrastructure that's out there that let's your application platform talk to other people's applications platforms. The phone switching fabric for the world is in there. The satellites, the X.25 pads, the routers. Everything is out here. Now, those are the four interfaces. Four interfaces. There are many others and one might say, "Are they not suitable for standardization?" We say, "Might very well be." These are the objectives that we have. These are the interfaces that are associated with; these are the objectives associated with the interfaces here. Let me give you an example of an interface which is eminently suitable for standardization but isn't showing up here. Back plane bus. BME bus, multi-bus, future bus - all those kinds of things that you'll slip neat cards into inside this application platform. But why isn't it on here? Why do we not want to open the box here? The reason is that

application portability is one of our primary objectives. If this application software knows what bus it is running on, that is not portable code, and you don't want to have that stuff hanging around, except in cases where you have to wring every last little bit of performance, or some other characteristic, out of this application platform. At the program level, you don't want to make a bus choice. You leave that to the people who are trying to get the job done.

One of our objectives is to get test technology in place to verify services and standards in each one of these interfaces. I think I may be retired before that happens. The reference model. Many people to whom I've talked about this being a Rorschach test see this as a technology model. They see technology when they use this. We talk a certain way to that kind of person. Other people see this as a programmatic description. This is a mapping of program responsibilities and a diagram that lets us identify who's responsible for what in a very clear and unambiguous way and negotiate the resolution with parties that we don't have any control over. One of the tools that a program can use is to cite the standards community in this sense. The POSIX 1003.0 has passed U.S. national ballot; it's now at the international ballot. You can point to this document and use the consensus that has been arrived at there, and the understanding that's been negotiated, and the vocabulary, and so forth where it's possible to do that, and you can really get your groups focused on the real issues.

This really is a program description and somewhere on your program. You probably need to have someone identifying your program's responsibilities in each one of these four areas. We've talked to people in programs who don't want to get involved in one or more of those specifications. Now, my group is the Distributed Systems Engineering Group, so we carry that one step further, and we talk about a complete distributed system. No new interfaces; no new entities. You've seen all this before, but this is what it looks like, and this diagram becomes very, very helpful in having discussions. When you say, this is our program, and by the way, these are that people we're having to deal with. These are the people completely outside of our control and you can begin to adjust and tune, you know, and get some credibility on your program as to how much you can drive these things and how much you need to respond to specifications on each one of these different areas. Now, there is a clear winner in each one of these areas. We expect that never to change; we never expect to see one solitary winner, there's always going to be. For instance, right now, in the APP, we have 35 standards here. Thirty-five individual standards that have been cited. One of those is always up for public review. Some of those are nearing their obsolescence stage. Some of those are brand new and haven't achieved widespread use, but there is a clear need there. We've had to pick a winner. So, there is always going to be a lot of life cycle. Think about it. Just 35 standards. We haven't seen a program yet that hasn't needed on the order of 100 standards. Managing that process and knowing where they're going and getting migration strategies in place and synchronized is a very difficult thing to do.

Let me give you the ones that we see as the center right now. Like I say, there is something in place. They don't work together well. But we are working on that. Let's see. Human-computer interface, IEEE-1295, used to be called Motif. Motif kind of won in the marketplace. Finally, all parties have agreed to that and IEEE-1295 is the API for human-computer interface. The style guide that's associated with IEEE-1295 is a document that was put out by OSF. In a considerably simplified version, it's also available from IEEE.

Information services, file systems, the POSIX .1 standard and POSIX .4 realtime file systems address this. SQL is for structured data. Okay, that is the center of consensus for structured data as opposed to file type data. Comp services — comp services here, the APP cites two. One points to the GOSIP specification. Another points to the ISDN specifications. We see both of those with the ISDN merging and evolving toward ATM.

Now, for the protocol side. We're also looking at a number of other interesting alternatives here. The NII is driving this area. There is a furious churn here with a lot of froth in the air. The power systems people, oddly enough, bring forth a very intriguing scenario. They say, "We've come to realize in running a wire out to everybody's house that we need to have an insulator down the center of that copper, and no reason not to use a fiber going down the center of that, a glass fiber, which we can then use for communication." You may not have noticed, but you have an option for reduced cost by letting the power company turn off your water heater once in a while. And that's the channel they propose to use to talk to your water heater. No reason you can't use it for movies on demand. There's an enormous bandwidth on that channel and more than any other carrier. By the way, what's their investment strategy for doing that? They're going to take the savings from their operational cost to pay for it all — no increases in rates. Gee, you know, think about what the cable TV people are proposing and what the Telco people are proposing. These guys are saying, "Oh, it won't cost you a dime." Think about it. There are some interesting possibilities over here that everyone needs to think about, and we expect this to accelerate. There are going to be more solutions out here as well - Com, API, Sockets. Go out there and pull some software down off the Internet — any kind of free software that does some interesting things. What's it written to? It's written to Sockets.

How many people here have heard of Mosaic? Geez, well, a lot of people. How many people have it running on their desk? Geez, okay. Good group, good group. Okay. Written to Sockets. You've got access! Oh, yeah! I forget which group I'm talking to here. Okay, I apologize. I should have assumed that. That gives you access to data scattered all the way across this planet. Okay. What's it written to? Sockets. Gee, are we going to rip that out and put something else in there. So, Com Services, Sockets, XTI (the other half of POSIX .12) is an important specification here. And finally, systems services. Systems services are those that manipulate things inside the application platform — event flags that start up new applications, and so forth and so on. POSIX .1 and POSIX .4 for realtime are clearly the winners in that area. That's where we stand. And that is a lot of functionality, and these things don't work together real well right now. X-windows, POSIX, and TCP/IP, you know, these things don't work together terribly well right now. There's a lot they don't do. Security is a myth in all of those areas. Management is, for the most part, a fond dream. But there are active communities under way right now that are trying to fix that even as we speak.

Last point. Program. This is the structure of our program and it's one that we recommend that groups look at to consider whether they've already got something going on that needs to be involved, or that they need to have in place to make their program work. We gather our OSE principles and guidance documents and you'll see, for each one of these labels, there is another backup slide, and I am not going to go through any of that kind of stuff. Just look for the labels on the top. We gather our guidance, this is available on our Mosaic server. I can

give you the URL off-line or, if you will give me a business card, we can get that back to you. The APP, the OSE guidance document. Gary Fisher is going to be speaking this afternoon on how we do procurement in this environment. We do other agency projects. We've worked with quite a few agencies on large projects, and we've learned a lot from them, and in a large sense, they're our laboratory, because a laboratory with a couple of computers running doesn't really give us the kind of laboratory that we need to get this work done. We work in standards forms. We have a distributive systems laboratory that checks out the technology, and we are trying to work out methods and principles for evaluating information technology, pieces of information technology, and characterize how well it works in the system in the large.

And finally the OSE conformance testing program. That's the testing program that right now gives you conformance testing and certification for POSIX .1 and is coordinating a number of other performance testing activities out at NIST and within the information technology community. My boss is very involved in that one, Roger Martin, and he sits on the Board of Directors of an activity in Japan. The European Commission and NIST in the States have funding to get a coordinative approach to testing with mutual recognition across multiple countries. One final word, and this is something I should have thrown up earlier in the presentation. Remember, I said services and services and an interface are your service requirements and that drives the selection of your standards.

Important programmatic principle profile. You create a profile for a specific need. And this is the point in the presentation when I should have put that up. What is a profile? Its just a citation of multiple standards for a particular use. Say you need to create a human computer interface profile. You'll create an API for programming; you'll create a style guide that tells you what happens with a human being; and you might create a protocol that runs out and says, "If there is an application that needs to interact with a human-computer interface over here, this is the protocol or message sequence that it is going to use." That's called the X-Protocol, by the way. It's from the X-Consortium. The one we have in here is X-Lad and IEEE-1295, which is the Motif API, and the style guide is the IEEE specification as well. That's a particular kind of profile. One of our biggest challenges is realizing that standards are very expensive. Profiles are maybe even more expensive, and what we are trying to do is to come up with a small set of profiles that we need to focus our energy on, and get a few profiles out there that have a lot, a lot of consensus. And gather people around that campfire maybe and let them then tailor those and explode out the diversity they need across the different programs. And we need the participation on multiple projects, real world projects. That's our methodology — to work with people who really have a real job to do and have real deadlines and deliverables and so forth and so on. Because we can always point to those folks when people say, "Oh, you can't do that in a standards committee." We say, "Well, they are going to ignore you if you don't." And people really need profiles right now. So, we've got a couple under way. There's one called a distributive platform profile, which is in final review in NIST. That answers the question, "What's the common subset for all different application platforms riding on a distributor system? What's the common subset that has to be defined and agreed to for all those platforms?" Well, we are coming up with something like that. Very ambitious. We hope not to get shot out of the saddle. But it will be an interesting experiment.

Another one is the infoserver profile that says, "Here is the platform that's sitting on the network and I am running on Mosaic server. What standards are involved in that? What specifications need to be put in place to allow people to access that data and for me to know what I need to put in place to make that possible for other people?" So, those are two profiles. We've identified hundreds, but we are trying to focus on those two and we are trying to get some folks in place who are willing to undertake experiments that we can help support or maybe that people can just do, and establish a campfire where people can bring them to the table and plop them down and we can all sit and talk about it. So, profiling is the final word of my presentation.

— NIST —

---

## Distributed Systems Engineering Group

Concepts and Methods for Developing  
Open System Environments

Fritz Schulz, Manager.  
Distributed Systems Engineering  
NIST/CSL

[fschulz@nist.gov](mailto:fschulz@nist.gov)  
(301) 975-2192

# Outline

---

OSE Program Context  
and Objectives



OSE Program  
Principles



OSE Program  
Elements and Status

# Distributed System Engineering Group

---

Facilitate the definition of Distributed Systems frameworks, architectures and methods required by organizations to define, operate, and evolve heterogeneous distributed systems. This includes refinement of the Open System Environment concept and continued development of associated supporting concepts, including:

- distributed system architecture principles and guidance,
- specific frameworks addressing user requirements, technical integration, and standards specification for distributed systems
- guidance on standards needed to satisfy specific objectives such as application portability and interoperability, DS Management, DS security, and internationalization.



# Open System Definitions

## Open System

- A system that implements sufficient open specifications for .  
interfaces, services, and supporting formats .  
to enable properly engineered applications software to:
- be ported with minimal changes across a wide range of systems, .
  - interoperate with other applications on local and remote systems, and.
  - interact with users in a style which facilitates user portability.

# **Open System Definitions**

---

## **Open Specifications**

Public specifications that are maintained by an open, public, consensus process to accommodate new technologies over time, and that, are consistent with international standards.

# **Open Systems Definitions**

---

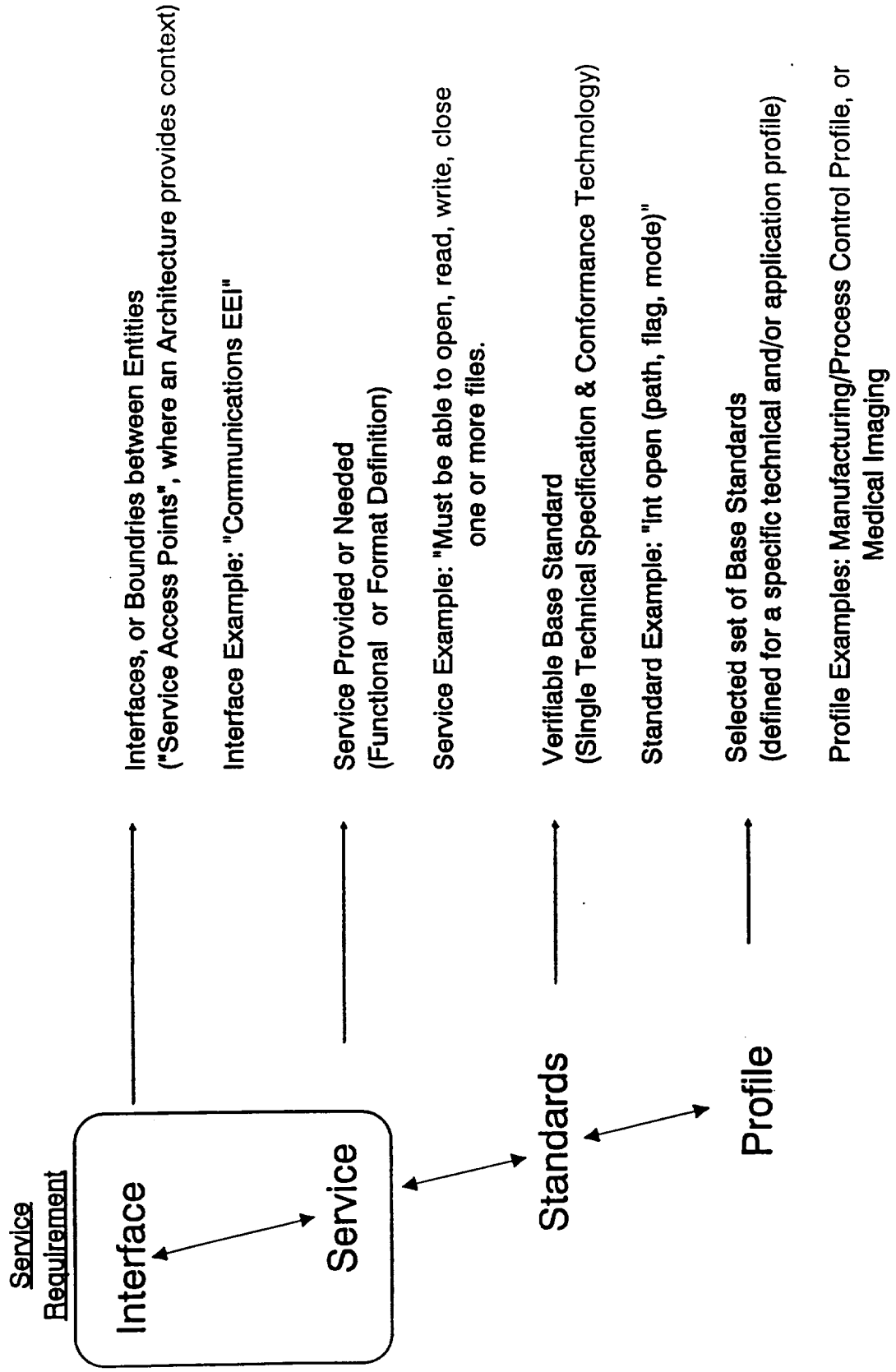
## **Portability**

The ease with which software can be transferred from one information system to another.

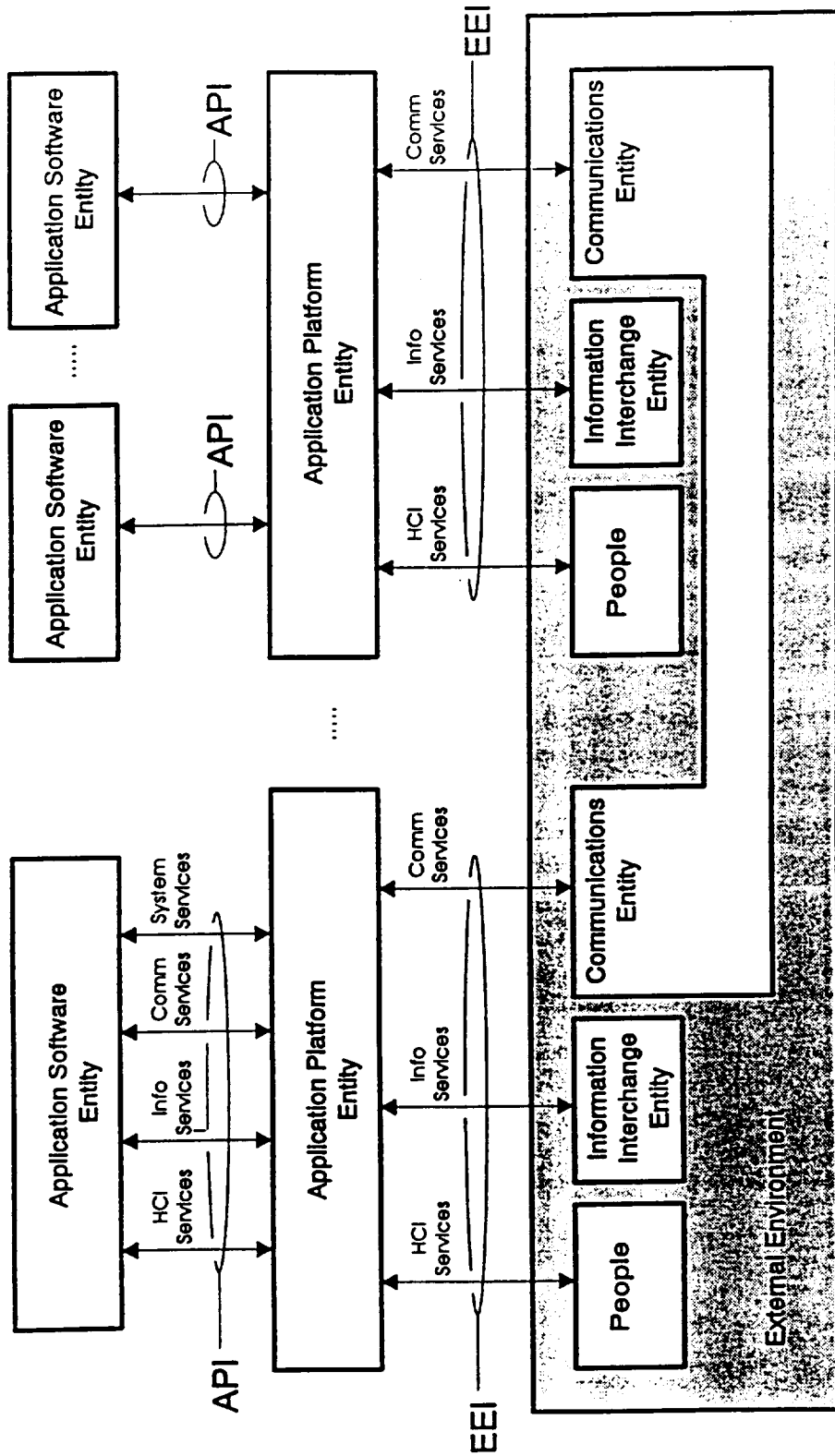
## **Interoperability**

The ability of two or more systems to exchange information and to mutually use the information that has been exchanged.

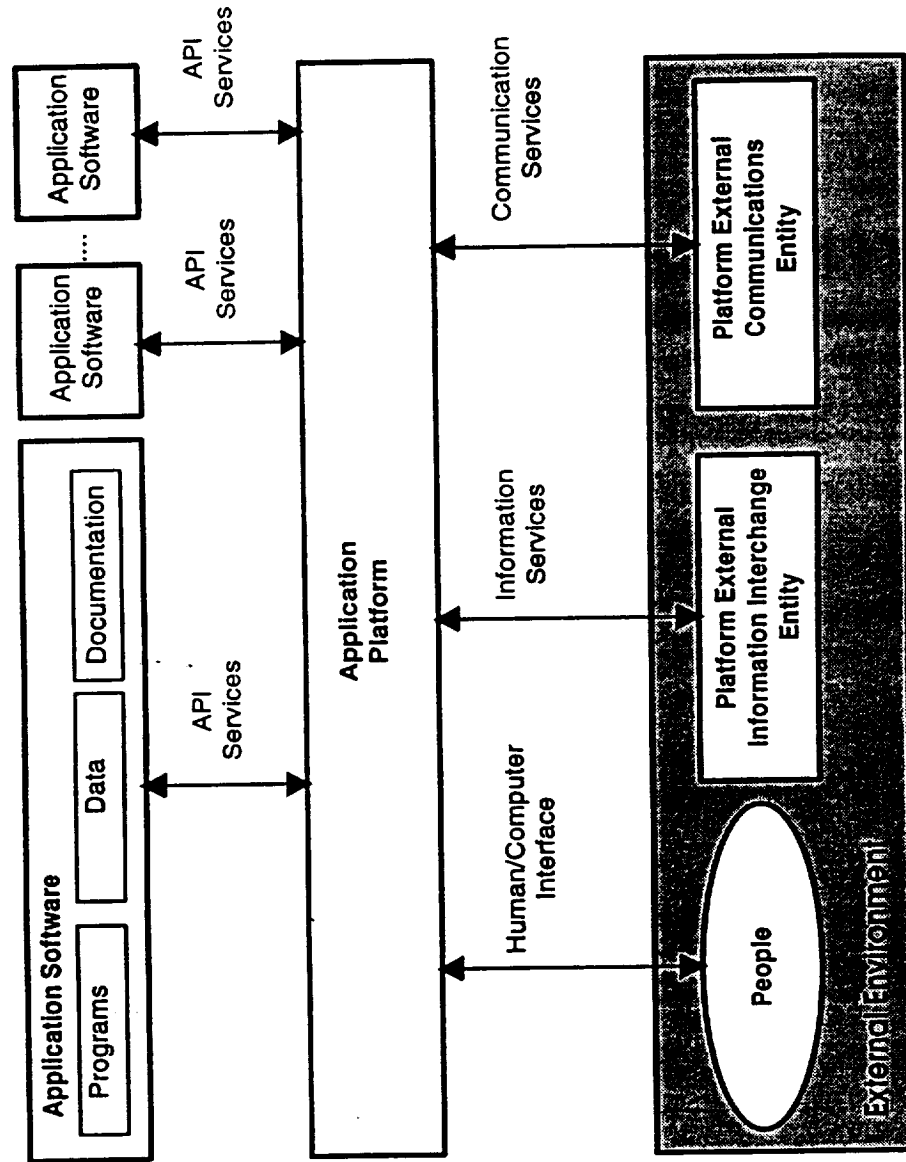
# Framework - ISSP Relationships



# Reference Model: A Complete Distributed System



# Reference Model: Entities

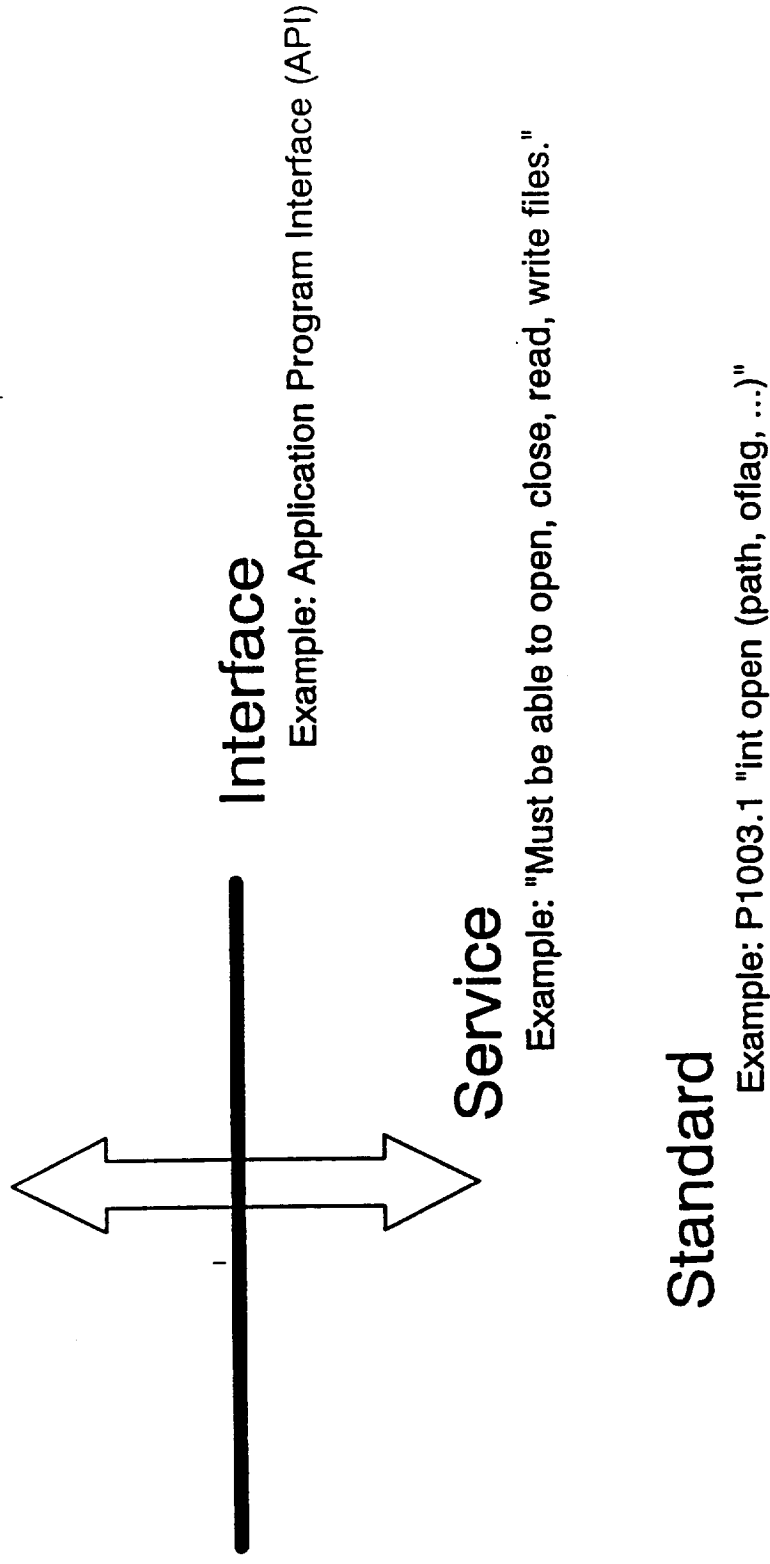


# DSE/OSE Program Overview

---

Other Agency Projects	Forums
OSE Principles  OSE Guidance Documents	Distributed Systems Laboratory - OSE Testbed -
	OSE Conformance Testing Program

# Requirement Based Approach to Open Systems





# OSE Technical Objectives

---

- Application Portability at the Source Code Level
- Application Interoperability
- User Portability
- Accommodation of Standards
- Accommodation of New Technology
- Application Platform and Distributed System Scalability
- Implementation Transparency
- Support Clear Statement of User Requirements

# OSE Program Approach

---

- 1) Support organizations participating in national class, OSE based, heterogeneous distributed systems by:
  - Establishing technical guidelines for participating systems
  - Identifying organizational best practice for making use of the technology and system
  - Establishing a Distributed System laboratory and procedure for evaluating candidate HDSE technologies
- 2) Provide a forum and process for establishing consensus among affected constituencies.

# Outline

---

OSE Program Context  
and Objectives



OSE Program  
Principles



OSE Program  
Elements and Status

# OSE Program - Distributed System Lab

---

- OSE Testbed is in operation
- Designed for Federated Operation with a wide variety of other participating testbeds
- Electronic Forum Support Experiment being used to provide a reality check for initial lab configuration, testing, and operation
- Laboratory Assessment Procedure is being defined, again driven by initial "trial" assessments to clarify objectives, effectiveness
- Trial HDSE "Components of Technology" assessments
  - OSF Distributed Computing Environment (focus on RPC)
  - Wide Area Information Service (WAIS) System and Z39.50
  - X-windows

# NII Principles and Objectives

---

- Promote private sector investment
- Extend the "Universal Service" at "affordable prices" concept
- Catalyze technological innovation and new applications
- Promote seamless, interactive, user-driven operation of the NII
- Ensure information security and network reliability
- Improve management of the radio spectrum
- Protect intellectual property rights
- Provide access to govt. information and improve govt. procurement

# OSE Program - Forums

---

- APP/OSE Forum (Federal Community Participation)
- POSIX P1003.0 - Guide to the Open System Environment
- OSE Implementors Forum (Profile Development)
- ISO/IEC JTC1 SGFS (Profiling)
- FIPS and Base Standards Committees, as needed

# OSE Program - Conformance Testing

---

- FIPS 151 (POSIX, ISO 9945-1) Conformance Test Suite
- Accredited POSIX Test Laboratories and Issue Certificates of Validation for Conforming products
- Assessing Test Technology for POSIX ISO 9945-2
- Coordinate with other NIST conformance testing activities
- Promote development of conformance testing to cover the full scope of OSE, and coordinate with related international test authorities
- Investigate Profile Conformance Testing concepts and technology

# OSE Program - Guidance Documents (1)

---

## Program and Process

- Principles for Establishing OSE Programs
- OSE Reference Model
- Distributed System Engineering Process
- IDEFO FIPS

## Issues

- OSE Procurement Guidance - Model RFP
- OSE Security Services and Guidelines
- Ops. and Admin. of OSE Distributed Systems
- Standards Selection Process
- Public Specifications



# OSE Program - Guidance Documents (2)

---

## Catalog of Base Strategic Standards

- Application Portability Profile
- IEEE Posix 1003.0
- ISO TR 10000-3 (OSE Profiling)
- OSE Implementors Workshop Documents

## Technology Profiles

- Distributed Platform Profile (DPP) -
- Distributed InfoServer Profile (DISP) -
- FIPS-158: X-windows - Distributed HCI -

## OSE Standards/FIPS

- FIPS 151-2 (Posix 9945-1) System Services API
- FIPS "N" (Posix 9945-2) Commands & Utils. HCI
- FIPS 158 (X-windows) User Interface API
- Posix P1003.12 (& FIPS) "Sockets" Network API

115

## Application Domain Profiles

- Medical Systems Architecture CRADA -
- Digital Library (supporting Dr. L. Welch) -
- Distributed S/W Development (supporting T. Rhodes) -
- Aerospace (Airframe and Manufacturing) -
- Military C4I (supporting Office of SecDef) -

# Open System Environment Procurement

Gary Fisher

We're going to talk about OSE Procurement. It's very easy to buy open systems (Viewgraph 1). How many believe that? All you have to do is know the insides and outs of about 30 different standards, and about 200 other specifications, and how they all relate to each other, and how to transition from what you have now to open systems. Well, fortunately, everything I'll tell you today is in a document we are getting ready to publish called *Guide on Open System Environment Procurements*. This is a general organization of that document (Viewgraph 2). I'll go over some of these topics here, the OSE requirements and specifications sections, and we'll hit on transition plans. The real benefit of this document is that it organizes lots of information that you wouldn't find anywhere else (or that if you did find it somewhere else, you wouldn't know how to relate it to anything else). We brought it all together. There are lots of lessons learned. We'll say a few words about what other organizations are doing.

Right now, what I would like to do is describe to you a little bit about what brought this document into being. We published the APP Guide, the *Application Portability Profile*, about two years ago (Viewgraph 3). Version one was NIST special publication 500-187; version two we modified somewhat and added some new specifications. That came out in June 1993. So, it's almost a year old. It's due for another overhaul, so we're going to make some changes, probably sometime this summer or early fall. Because of the application portability profile and the open system environment that it describes, people are buying and building open systems.

That's kind of a misnomer when you say, "I want to buy an open system." What you want to do is establish an open system environment. We want applications that are affordable, scalable, and interoperable across a broad range of platforms and computing environments from very small microcomputer desktop machines to very large supercomputer processing mainframes. And we want to do that in an environment of networking where we essentially have anybody's machine connected to anybody else's machine. We get our applications to run on anybody's machine, using anybody's network, using anybody's database, using anybody's operating system, and they all run the same. That's what we're really looking for. Unfortunately, we can't do all of that right now, but we are getting closer and closer as the standards develop. Building an open system environment is complex, a long-term project. We're talking on the order of five to ten years, generally, for a large organization— for the Department of Defense, probably ten years. Maybe one of the services, a smaller service— Coast Guard, Air Force— maybe five years. NASA. Who knows? That's for you to determine. There are lots of lessons learned for the simple fact that there are procurements on-going; there are procurements that have been completed; billions of dollars worth of open system environment.

Infrastructure is in place; it is being put in place; hundreds of thousands of users are already affected. There is a redeeming factor about open systems: you don't have to do it all at once. You can do it a piece at a time. This is the transition part of the procurement guide (Viewgraph 4). In the scope, we debated long and hard who we were really directing this document to, and we decided at the end that it has to be a fairly abstract document. It can't contain all the information that's in the standards, of course. I mean, you'd wind up with a document at least three feet tall.

Program managers and senior project engineers: they're the people who need to know the information in this document. We provide a sort of decision model. It's not really a decision model, but you will find all kinds of decision points in the report, and we will give you all kinds of information. What happens when you go this way? Why you should go that way? We also give you lots of guidance on the applicability of specifications, not only in where they apply in a particular application environment or an operating environment, but also when you choose this, what else applies. When you choose one specification, what else do you have to look at? The lessons learned we provide to assist in the decision making process. Generally speaking, the lessons learned are fairly easy to identify because they're in dark colored boxes that say "lessons learned" at the top. There are other boxes that are not lessons learned, but they contain lots of important information. And of course, we assume familiarity with the OSE in the Application Portability Profile. This is how we organize everything in the guide (Viewgraph 5). We talk about the relationship of the OSE to the RFP process, then we go on to the individual specifications in the OSE service areas. They parallel what's in the Application Portability Profile. We talk about standards testing, validation interoperability testing. Lots of people don't understand testing, and when they require certain things, they find out they get what they ask for and it's not exactly what they wanted.

We include organizational requirements in some instances: in each section, you'll find subsections, for instance, that provide information to contractors or information to the people who are writing the RFP and what to tell a contractor so proposals can be evaluated. And we also tell you, in many instances, how to evaluate the proposal, what to look for, and the responses to expect. Briefly, here's how the RFP process and the OSE relate to each other (Viewgraph 6). You start out, of course, with organizational mission requirements. Those generate information technology requirements. Some of them can be met in terms of open systems.

Now, part of the job is determining which ones are the ones that are open system requirements and which ones aren't. If you have a requirement you think is an open system requirement, you might go to the guide on open systems procurements and see if it really fits as an OSE requirement. If you still have questions, or you find out that it does apply, then you go to the experts on those standards. You can't do this by yourself. You have to consult with experts. There's too much information to worry about. There are too many ways you can hang yourself if you don't have expertise available to help you get the job done. In the procurement guide, there's actual RFP text that you can insert in an RFP, depending on how you decide to use that particular requirement. And there are also evaluation factors that go along with the RFP text. A request for proposals is issued. The proposals, as they come in,

are evaluated and an award is made. Implementation is then taken care of. What we are trying to do is assist those folks who are writing the RFP to make sure that they are writing in the right terms and they are asking for what they really need.

This is an outline of what we recommend you should actually put in the RFP (Viewgraph 7). There's a section on requirements for open system environment; that's where we tell the contractors or the proposal people who are going to submit proposals that this is an open system environment procurement. Then we go through each of the APP service areas, talking about operating systems, human computer interface, and so on and so forth - graphics, network and security and management services. There are sections in the report for each of the services, each application, communications, and other requirements (Viewgraph 8).

What you're doing is trying to take existing legacy systems and convert them to open systems, or you're buying new systems which you want to operate in an open system environment. Therefore, all of these systems and applications have to fit within that envelope. You can talk about local area networks, wide area networks. Legacy systems. Everybody asks, "What do we do with the ones that we're not going to transition? They're going to go away sooner or later. And they're not going to be around for an open systems environment. What do we do?" Well, here is where we talk about those individual systems, what interoperability is required, how we're going to share data, or what the requirement is to share data. You're waiting for the vendors to come back and tell you how they're going to do that after they have looked at the systems. And then, of course, there are organizational requirements consisting of who the users are, where they are, the number of locations, and organization responsibilities. I am going to skip this next one for the time being (Viewgraph 9). I'll come back to it in a few minutes.

As I said, we also get into standards testing (Viewgraph 10). You have no idea what I went through to gather this information. And I work at NIST, which does the conformance and validation testing. When we talk about validation, we're testing conformance of an implementation to a standard. When we say "conformance," we're saying, "How well does it meet the requirements in that standard?" Validation says we only tested it. It either passed or it failed. On interoperability - we're testing communications, generally speaking, but we're also talking about data sharing or the interchange of data. All the vendors gripe when we talk about validation. They say, "It's gonna cost an arm and a leg to get it done." Well, this is what it costs. Anywhere from \$2,500 to \$100,000 per implementation, depending on what you are trying to get validated. Say an SQL implementation costs around \$15,000. A GOSIP implementation, depending on what has to be tested, costs anywhere between \$20,000-\$100,000. Communications testing is very expensive, of course, because there are only a few people who do it, and they own the market. The demand is also dependent on how many people are in the pipelines, the queues for getting tested. How many accredited laboratories are available to do the testing? It all depends on how much NIST has to get involved with the vendors themselves to get them to pass the test, how much time we have to spend on their sites, and so forth. There are all kinds of fees involved with this. This doesn't even include the fees that are associated with having a third-party test laboratory do the testing.

We will go over types of validation, delayed validation, prior validation testing, and prior validation. Everybody misunderstands these terms. They're in the GSA, ADP and Telecommunications Standards Index. Delayed validation: we have a closing date approaching on an RFP. We know there are no implementations right now that are validated. We're going to allow people, after we've gone through the proposal process, maybe even after we've done the award, to get validated at that time. Prior validation testing says we may not have enough implementations right now to have a valid procurement. It won't get us the best choices. So, what we need to see from the vendors is, "Yeah, we have either implementations already tested (they're not exactly what you want), but we have other implementations in the pipeline. We're going to get them tested." And they can usually prove that to you by showing you a contract or a letter of intent from the testing laboratory. Prior validation: we won't accept anything in a proposal except validated products. They have to be tested before you submit your proposal.

There are different classes of validation. There is base validation, derived/registration validation, and demonstration. Base validation says that this is the implementation, this is the platform we put together and tested, and that's what gets listed on the certificate. Derived validation, or what Ada calls registration: "Here is a certificate. We took that implementation and put it on this other machine, ran the test against it, but we didn't have government witnesses and we didn't go through the process. But, take our word for it, it passed the test." That's a derived validation. NIST will list it, but we won't issue a certificate.

Conformance demonstration: either one of several possible situations has arisen here. There's no Federal Information Processing Standard (FIPS). On the other hand, maybe a FIPS exists. But we don't have a test suite for it, or we don't have an accredited test lab or there's not a test procedure. One or more of the parts is missing. So, what we say is, "Okay, show us that it works." It's what you call FCD, Functional Capability Demonstration, in procurement parlance. It takes the place of testing in that respect.

Every group within NIST has a different way of testing, because all the standards are different. It takes a lot to understand. How much testing is enough? (Viewgraph 11) The answer to that question is how much risk are you willing to assume? How much do you want to pay for it? You could accept the manufacturer's declaration: "Yeah, we tested it. It works. Trust us." That's the highest risk, lowest cost. Is it really the lowest cost? You don't know until you try to implement. Right? If it sets you back a month in your implantation plan, is it really the lowest cost? Derived validation (and this is the situation I just described with Ada) happens with other things too, like SQL and some of the more esoteric standards. There's high risk and low cost there. It's already validated somewhere. And we're just accepting it on a different platform. Validation on the proposed platform entails intermediate risk and cost. "We've already had it validated, and it's the one we're bidding to you on this procurement," says the vendor. We accept it that way.

And then there is product suite validation on a proposed platform which is the least risk, highest cost. And what this says is, "Mr. Vendor, you are going to give us an SQL implementation, a C Compiler, an Operating System Interface, X-Windows, GOSIP, etc. We

want them all tested on that platform that you are going to bid, and we want to see that they have all been tested for interoperability. In other words, we're running GOSIP with the Posix implementation, with the SQL implantation, with the compiler." The most expensive way you can test it, but I know people who did that. The United States Army did that in their last buy - Sustaining Base Information Services (SBIS).

This is for one platform. Now let's multiply the cost by the number of platforms. I think you can get an estimate of which one of these is going to cost you, and which ones are going to give you the lowest risk. If, for instance, it costs \$150,000 to get one platform, a complete suite, of software tested, and let's say there are five different platforms, that's \$750,000 right off the bat you can expect to pay. It might not be a line item, but the cost is in there somewhere. There are alternative standards testing. We're really going after portability, scalability and interoperability (Viewgraph 12).

On the next slide is interoperability (Viewgraph 13). These are just some of the options that you might want to try in deference to paying for standards testing. But when you want to compile and execute a selected test program on a proposed platform, you have to make very sure that it is going to work. So you have to have your own internal expertise in testing, to sit down and go through the process of trying to get the product to work, or to make sure the program compiles on the machines that you want it to compile on. You know there are no extensions in it. You don't want the vendor to have to come back to you and say, "This program doesn't work because you used a non-standard compiler."

Scalability: it involves the same type of concept. You're just moving the program from one machine, one architecture, to another. And generally speaking, we're talking about going across vendor lines.

Interoperability: a very simple way to do this is just to transmit a file from one machine to another machine, using the communications protocols that the vendor is proposing. And then send it back. Then do a file compare and see if they're the same files. It's the same concept with electronic mail messaging and binary files. Start off with an ASCII file, but then wind up with binary files and messages. There are different ways of skinning this cat. It all depends on what you're willing to put up with and how much you're willing to invest in the procurement process itself.

When I said we have RFP text available within the guide, this is what I was talking about (Viewgraph 14). Generally speaking, all the text in the guide, the informative text, is normal Times Roman type font. Anywhere you see italics, that's where we're talking about RFP text. The report will be electronically available, so you might want to just edit it right into a document. But this is the type of text we're talking about that we'll provide for you. It's been used in other procurements; it's been modified; it came from GSA in many cases. There's a lot of work that has already gone into the text. Here's an idea of how you would write a validation requirement, for example (Viewgraph 15). This is just an idea of what conformance demonstration says (Viewgraph 16). There is clear-cut text that you can throw into the RFP.

This is where it starts getting interesting. In each of these sections, we have included subsections, one of which is "instructions to the contractors" (Viewgraph 17). This is what you tell the contractors. You can get very definite when it comes to open systems, because we have seen a lot of what vendors have to say about open systems. And not everything we'll like. One must be able to tell the difference between open systems and open systems marketing. Vendors are masters at open systems marketing. So, you have to get rid of the chaff somehow, and one way of doing that is to say, "Okay, you can give us all the marketing literature you want to, but we want to see your validation certificates; we want to see the test results, summary reports; we would like to see a script for your conformance demonstration; if you have some alternative specifications you want to use, give us the reasoning behind recommending those alternatives."

You want to see a cost/benefit risk analysis. Vendors don't like taking any of the risk of putting together a proposal for an open system environment because it bares their souls. You get to see straight into their hearts. Everybody is playing with the same sheet of music, the same standard. The difference from one vendor to the next is, if they're doing the same things, there is a cost/performance tradeoff. Get the fastest machines for the lowest cost at that point, because they all do the same tasks.

Here's a lesson learned (Viewgraph 17). That's what the box looks like. That's what you look for when you see the document. Along with the instructions to the contractor, we talk about evaluation of proposals. And for each of these sections, you'll find a section like this, a subsection: "For each GOSIP protocol stack submitted, registration should be indicated" (Viewgraph 18). And there are some other forms. One of these forms I skipped over a while ago is one of the means of keeping track of some of this information (Viewgraph 9). It's a simpleminded way of handling information, but it turns out that if you have that information all in one place, it becomes a very simple task to determine whether they meet the requirement. If you don't, you search for weeks trying to find the information. That was a hard lesson learned. Along with all the different service areas that we talked about in the document, we also talk about standards profiles. Fortunately, everything having to do with standards falls into the OSE standard profile (Viewgraph 19). But you can also give the vendors the chance to come up with recommendations for other specifications that can be used in concert with, or as complements to, the OSE standards. This is where you tell them to give you the rationale for their use - a cost/benefit risk analysis. You especially want to know the effects of a specification's use on transitioning to OSE. It might not be a good idea to use these alternative specifications.

We haven't forgot hardware requirements (Viewgraph 20). Reviewers of the report asked us, "Please put something for hardware requirements in there." This is what we added. These are all basically items for which there are standards or for which there are known requirements that everyone uses as a rule-of-thumb. Accessibility requirements were not forgotten. There are people who don't have the same capabilities as a lot of other people do, such as those who have a disability in walking, or are blind, or maybe deaf. We also talk about use of government-owned equipment in here, and of course, telephone and other system requirements and hardware constraints that we already know about.

Here's the difficult part of putting together an RFP. Do you want the contractor to perform the transition for you, or are you going to implement some way of controlling transition over the systems? If you decide one way or the other, you're probably going to have to go through this process (Viewgraph 21).

Plans and strategies for transition. You have to build a baseline definition - do an analysis of it (Viewgraph 22). You have to know where you are to determine where you're going. And in a lot of cases, agencies don't know where they are right now even with what they have - even if it is a closed, proprietary solution to their information system needs. The Army did a study to find out how many actual applications and systems there were in the Army inventory. They concentrated on the number of applications and the languages and the users and interfaces to other systems. And those interfaces also included data that was exchanged. They found out they had something like 3,300 applications throughout the Army inventory. These were administrative systems. They got rid of the duplicates and merged other ones, and they found that they only needed 1,500 of those applications that were unique, that the functionality wasn't accomplished anywhere else. They found seventy-some thousand data elements, and they went through the process of eliminating duplications and came up with 12,000 that they actually needed. By just going through this process of trying to find out where everything was, they eliminated over half of their information processing requirements and three-quarters of their data requirements.

We develop an objective architecture. This is really what we're talking about: this is the way we see the open system environment five years, ten years down the road. This is what we're building to, the direction we are going, and then we implement the intermediate targets (Viewgraph 22). Here's where we are; here's where we're going. We don't have to do it all at one time. Remember, I said we can do it in stages. Well, those intermediate targets are those stages. This is just a graphical representation of what I mean. The further along the plan we move, the fewer proprietary systems we have in place, the more open systems we have in place to we reach that final objective. That we never reach that final objective is the trouble. As we move along, open systems continue to evolve. New technology comes along that we want to plug in. If there are standards available, and they meet our other requirements, that pushes the objective out a little bit further. And we can get rid of some of the things that we're doing back here. Target one may go away after five years. You're already past it, but that technology is old. You've got technology out here now that is new to replace it.

Transition strategies are the kind of guideposts that people need to make decisions. They come to a fork in the road and they say, "Well, which side do we take?" You look at the transition strategies and you find one that fits your situation and you say, "Well, that says we have to go this way." That's how you decide. We may want to integrate CASE technology where we don't have it. Maybe we want to have centralized data standardization. We don't want to let all the operating units decide how they're going to do things. We're going to control that centrally. Maybe we'll let them handle their own data updates. Maybe we'll decentralize transaction processing. We'll go to a client-server architecture. We're going to buy off-the-shelf rather than build systems. And everywhere we want graphical user interfaces. We don't have that now. We're going to go to X-Windows.



These are all strategic directions that I am really talking about here. And they help the people who are trying to write procurements and make and do the evaluations and determine what they should be doing. In developing an objective architecture, you might start with something like this (Viewgraph 26). You would probably find this in a baseline document. If a contractor is doing it for you, what they are going to do is come back and say, "You have Building A over here and you've got Building B over here. This is what's in Building A and this is what's in Building B; this is how they are connected and this is what the machines are." You will see a lot more detail. This is a very high-level schematic. You'll see a lot more detail when they talk about the individual products that operate on those different platforms and the applications and where their data bases are stored. You see that information when you get down to a lower level. And what they'll do in the objective architecture is say, "Okay, we're going to apply all these standards and we're going to get rid of some of these systems. We're going to add these other plans and systems. We're going to connect up this extra building out here. We're going to go on line over here," and so on and so forth. You'll see changes develop in this schematic that reflect the decisions to do those different things.

Now, for each transition, for each intermediate target during transition, you may find a different one of these diagrams. It just changes a little bit from the previous one. Instead of having Ethernet hooked in here, it may be replaced by TCP/IP to bring you up to a certain level of functionality. And then the next diagram may include some GOSIP protocols or some GOSIP routing, and then further on, another change occurs. Maybe this mainframe goes away. That's what you are looking for. The objective architecture is defined as a kind of transition concept. You don't have to define it 100% right now. Here's another one of those forms I was talking about that simplifies life for you (Viewgraph 25). When you're trying to evaluate proposals, you want to see what the vendor has done for each one of these different service areas and the different platforms that they're proposing. You can slice this information different ways, but this form turns out to be one of the most effective ways because vendors have to put a product in for each one of the boxes. If you find empty boxes, you start looking at them and you say, "Um, I wonder if they didn't know what that meant or if there aren't any products?" Then you start digging and you find out. You'd be surprised at how much the vendors learn going through one of these procurements.

Something that is developing and has come up over the last year and a half or so is this concept called middleware. What the vendors do is say, "We're going to write all your applications to our middleware. And we're not going to just stop at places where there are proprietary hooks to the application. We're going to do it from the standards side, too." And where there are standard interfaces? "What we'll do is to write all of the applications to our middleware, and then we'll do the translation to whether it is standard or non-standard." Not good. Here we have a standard. Why not write the application to the standard. Why write it to somebody else's concept of what the API should be? You're hooked into these people now. It's the same thing as not having any standards at all in place. You may need, in certain cases, some middleware. For instance, right now we don't have any standard implementations of X-Windows. We do have X-Windows systems. MOTIF is a user interface that a lot of applications and developers like to use, but unfortunately, it's proprietary. But people use it anyway. So why not write to a MOTIF middleware piece and then, when the standard

develops, we can replace that piece of it with the direct hooks to the 1201 standard (IEEE P1201.1).

There are several annexes (Viewgraph 27). Annex A provides general information about evaluation strategies, structures, that we're proposing. We give an example of evaluation factors down to great detail. We don't go through the whole process, but they are very detailed. In Annex B, we've included an example statement of work, based on a procurement for office automation. It's just for illustration. You may not agree with it, but all the text is taken out of the procurement guide. And, of course, there are references, a glossary, and an index (Viewgraph 28).

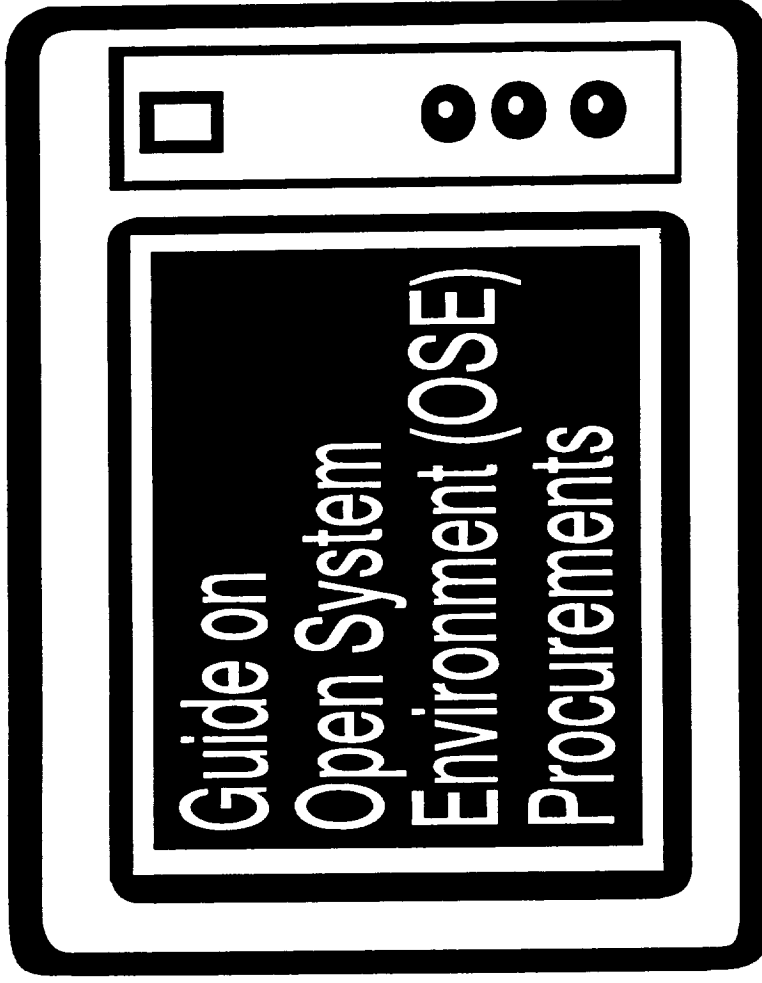
We parallel in the evaluation strategy and the APP so that you're looking at groups of services that everybody would be familiar with. Of course, we need to evaluate the total management technology and a cost profile for each contractor, and we say a few words about that. The evaluation of the transition to the open system becomes very important. Understanding of the environment's complexity, knowledge of the OSE and standards, planning and scheduling realism on the part of the vendor in the proposal is what you really need to dig into. That's where you make the determination about whether the vendors really know what they are doing. This is kind of an overview of the evaluation - the Source Selection Evaluation Board (Viewgraph 29). There is an OSE team up here, which is really a technical support team (Viewgraph 30). What they do is to look into all the different matters that seem to pop up when you are going through a major procurement that has to do with open systems - checking out validations, checking out whether the platform is actually commercially available, etc.

I will skip back to examples of evaluation factors. Like I said, we get down to real detail levels here. There are just a few for each of the services that are in the report (Viewgraph 31). Then when you get back in the Annex B, this is what you will see (Viewgraph 32). These paragraph numbers say that this is a second level paragraph, this is a third level paragraph, you can indent it when you're putting together your RFPs. But the text is there for you to use as an example. I covered everything that is in the document. If you have any questions about a particular subject or topic, I would be glad to go over it. We're hoping it's going to be a best-seller, particularly for the program managers and software engineering/systems engineering types. The caveat is this: If you're buying open systems, or you're buying services from a contractor, or buying implementations of information/infrastructure to support other systems, you've got to get this in-depth expertise on software development, communications, and database technology. And you've got to have these people know what the standards are about and how they fit together. You make the decisions of where you're going with all this and they have to implement. You can only be prepared (Viewgraphs 33 and 34).

# ***OSE PROCUREMENT***

---

**Gary E. Fisher  
Computer Scientist**



**National Institute of Standards  
Building 225, Room B266  
Gaithersburg, MD 20899 USA**

**Phone: (301) 975-3275  
PHAX: (301) 926-3696  
Internet: [gfisher@nist.gov](mailto:gfisher@nist.gov)**

# ***OSE PROCUREMENT***

---

## **Introduction**

- **Background**
- **Purpose and Scope**
- **Organization of the Guide**
- **OSE Requirements and Specifications**
- **Transition Plans**
- **Annexes**

# ***OSE PROCUREMENT***

---

## **Background**

**NIST APP Guide (NIST SP 500-187, and 500-210)**

**Agencies are building OSE**

**Building an OSE--complex and long-term**

**Lots of lessons learned**

**Redeeming factor--OSE not done all at once!  
Can be built in stages.**

## ***OSE PROCUREMENT***

---

### **Purpose and Scope**

- Audience--Program Managers and Senior Project Engineers
- Model and guidance for planning and specifying OSE procurements
- Guidance on applicability of specifications
- Lessons learned--assist in decision-making
- Assumes familiarity with OSE and APP

# ***OSE PROCUREMENT***

---

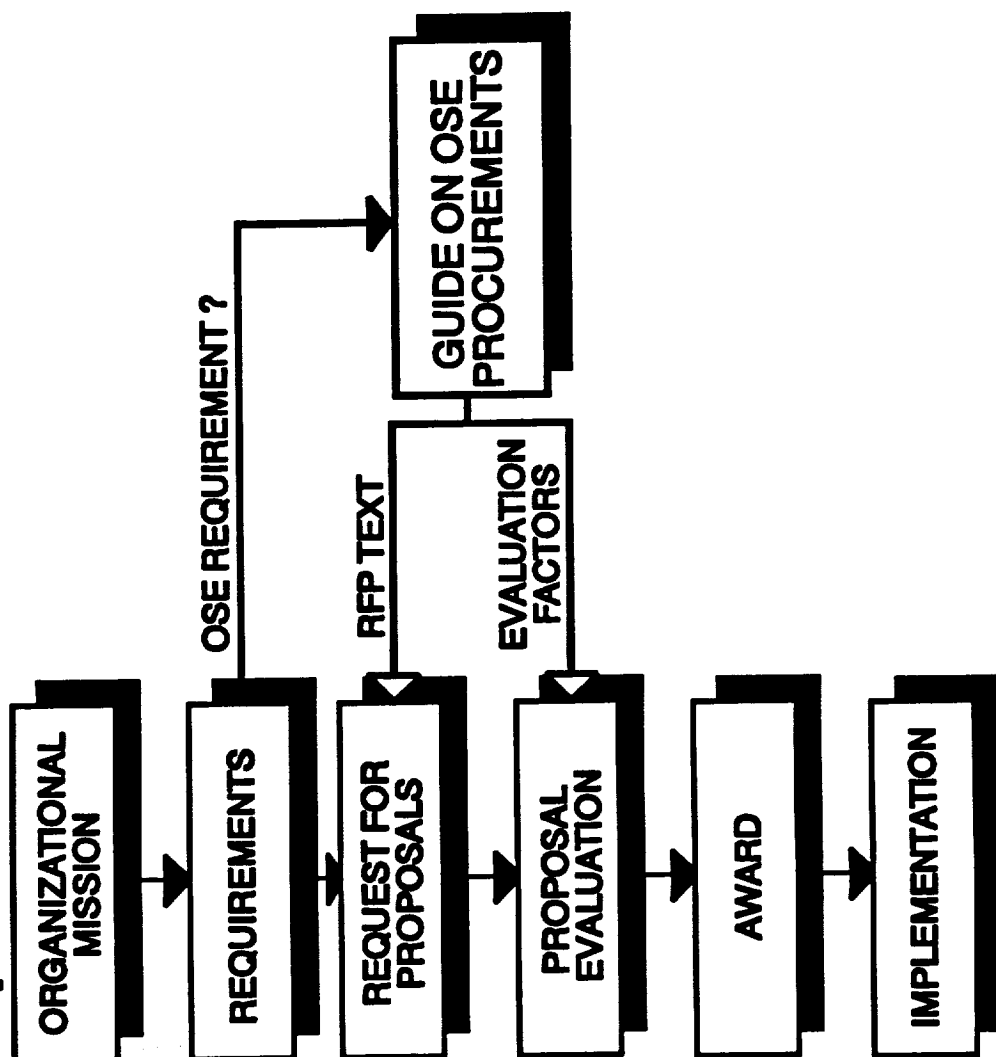
## **Organization of the Guide**

- **Relationship of OSE to RFP process**
- **Specification of individual OSE services**
- **Standards Testing (validation and interoperability)**
- **Organizational Requirements--Including Instructions to Contractors, Evaluating Proposals**

# ***OSE PROCUREMENT***

---

## **Relationship of OSE to RFP Process**





# ***OSE PROCUREMENT***

---

## **Specification of individual OSE services**

- **Requirement for Open System Environment**
- **Operating System Services**
- **Human/computer Interface Services**
- **Software Engineering Services**
- **Data Management Services**
- **Data Interchange Services**
- **Graphics Services**
- **Network Services**
- **Security Services**
- **Management Services**

## ***OSE PROCUREMENT***

---

### **Requirement for Open System Environment**

- **Systems and Applications--Payroll, Time and Attendance, etc.**
- **Local Area and Wide Area Networks--Existing and planned**
- **Legacy Systems--Non-transitional systems, interoperability, shared data**
- **Organizational Requirements--Numbers and types of users, locations, organizations**

# OSE PROCUREMENT

## OSE Configuration Summary

CLIN	PRODUCT NAME	PRODUCT IDENT.	PLATFORM	VALIDATION TYPE	VALIDATION IDENT.	DESCRIPTION	REF. DOC.
0100	ABC-SQL	Release 1.01	486/100	VPL	NIST/930929-0001 VPL 6/93, p. 37	Data Management, ABC Company	Vol. 1 #105

## ***OSE PROCUREMENT***

---

### **Standards Testing**

- **Validation--test conformance to standard**
- **Interoperability--test communications**
- **Cost**
  - **\$4,000 POSIX**
  - **\$11,000 compilers**
  - **\$15,000 SQL**
  - **\$20K to \$100K GOSIP**
- **Types of validation**
  - **Delayed validation**
  - **Prior validation testing**
  - **Prior validation**
- **Classes of validation**
  - **Base, derived/registration, demonstration**

# ***OSE PROCUREMENT***

---

## **Standards Testing Guidance**

- **How much testing is enough?**
  - **How much risk is reasonable?**
- **Levels of validation risk assumption**
  - **Manufacturer's declaration--highest risk/lowest cost**
  - **Derived validation--high risk/low cost**
  - **Validation on proposed platform--intermediate risk and cost**
  - **Product suite validation on proposed platform--least risk/highest cost**
- **Multiply by the number of platforms to get estimate of standards testing cost**

## ***OSE PROCUREMENT***

---

### **Alternative Standards Testing**

- **Portability Conformance Demonstration**
  - **Compile and execute selected test programs on proposed platforms**
  - **Validated Product List (VPL) entries for two or more platforms**
  - **Trademarking or branding certificates and test results**
- **Scalability Conformance Demonstration**
  - **Compile and execute selected test programs on proposed platform**
  - **Repeat on platform of different architecture**

# ***OSE PROCUREMENT***

---

## **Alternative Standards Testing**

- **Interoperability Conformance Demonstration**
  - **Transmit file from one platform to another**
  - **Transmit and receive electronic mail message and binary file between platforms**

## **OSE PROCUREMENT**

---

### **Operating System Services**

***The contractor shall provide operating system services including the specified interfaces, protocols, and supporting data formats for implementing portable applications at the application-operating system interface. These services shall include kernel operations, commands and utilities, system management, and operating system security as prescribed in the following:***



# **OSE PROCUREMENT**

---

## **Validation Requirement**

***POSIX-like operating system environments for kernel operations offered as a result of this and other requirements in this contract shall implement FIPS PUB 151-2, as a minimum, and shall require validation in accordance with provisions contained in FIPS PUB 151-2.***

## ***OSE PROCUREMENT***

---

### **Conformance Demonstration Requirement**

***Electronic Data Interchange (EDI) services offered as a result of these and other requirements in this contract shall conform to FIPS PUB 161-1 EDI and shall require conformance demonstration.***

# **OSE PROCUREMENT**

---

## **10.1 Instructions to Contractors**

**Contractors should be required to submit registration of GOSIP products. Not only should software implementations be registered, but any hardware components submitted with integral communications software, such as gateway platforms, should be registered. The acquiring agency should check the current GOSIP registration database for registered means-of-test (MOT) and abstract test suites (ATS) used. Often, a GOSIP MOT is not available for testing a particular protocol, in...**

**Lesson learned...  
Note that GOSIP registration testing is particularly expensive in comparison to other validation testing.  
Conformance demonstration for cases where GOSIP tests do...**

## ***OSE PROCUREMENT***

---

### **10.2 Evaluation of Proposals**

**For each GOSIP protocol stack submitted, registration should be indicated. Each attachment to the network, such as gateways, network printers with embedded communications, etc., should be accompanied by VPL registration. Additional information about GOSIP procurement may be found in CSL Technical Report CSL/SNA-93-3, "Procurement of U.S. GOSIP Products--Advice to Agencies.**

# ***OSE PROCUREMENT***

---

## **Standards Profiles**

- **OSE standards (FIPS, IEEE, ANSI, ISO, others)**
- **Other specifications (contractor recommendations)**
  - **Rationale for use**
  - **Cost/benefit/risk analysis**
  - **Effects of use on transition to OS**

# ***OSE PROCUREMENT***

---

## **Hardware Requirements**

- **Monitors, printers, input-output devices**
- **Accessibility requirements**
- **Use of Government-owned equipment**
- **Telephone and other system hardware constraints**

# ***OSE PROCUREMENT***

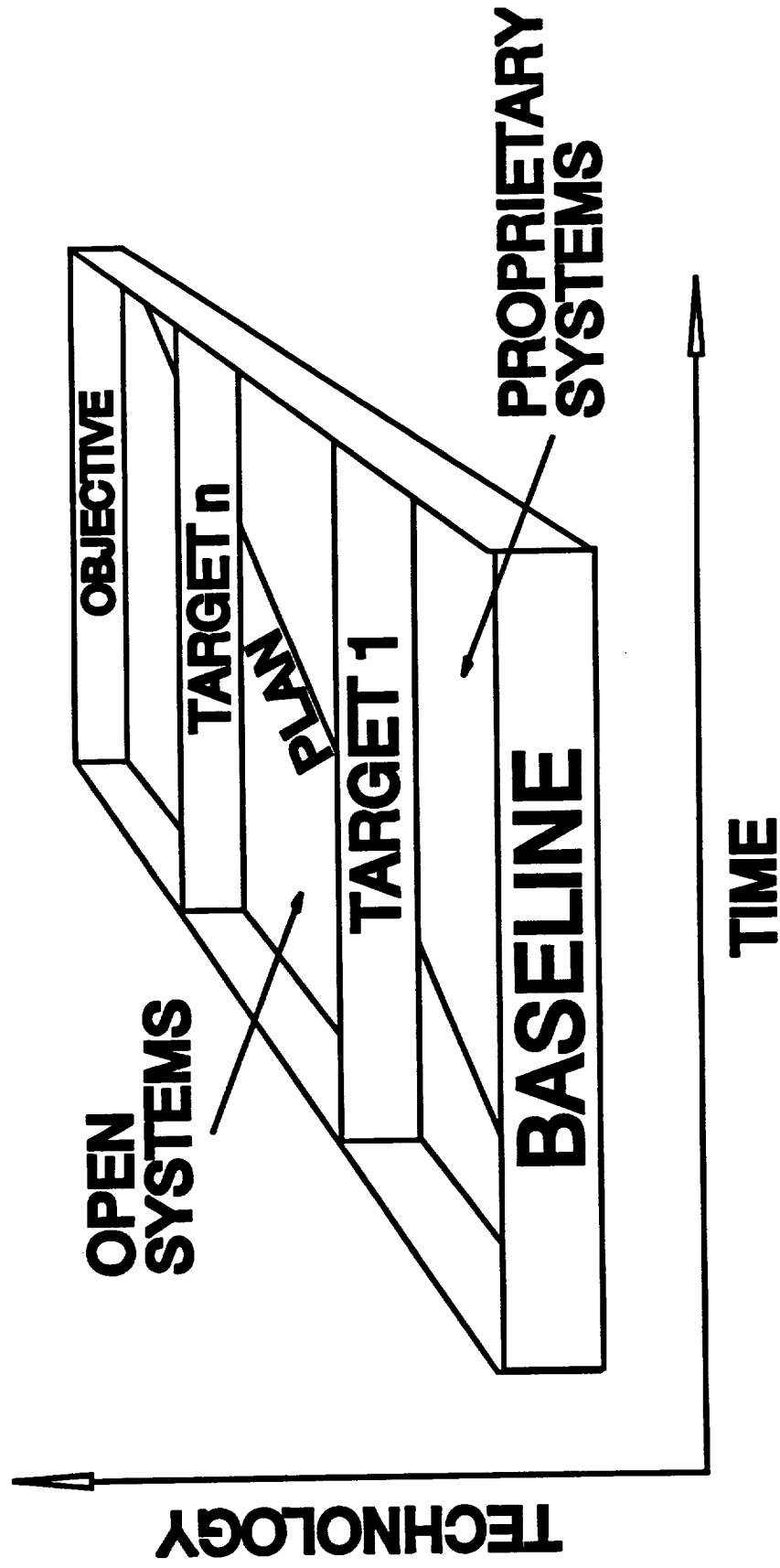
---

## **Transition to OSE**

- **Plans and strategies**
- **Baseline definition and analysis**
- **Develop objective architecture**
- **Implement intermediate targets**

# **OSE PROCUREMENT**

## **OSE Transition Framework**





# ***OSE PROCUREMENT***

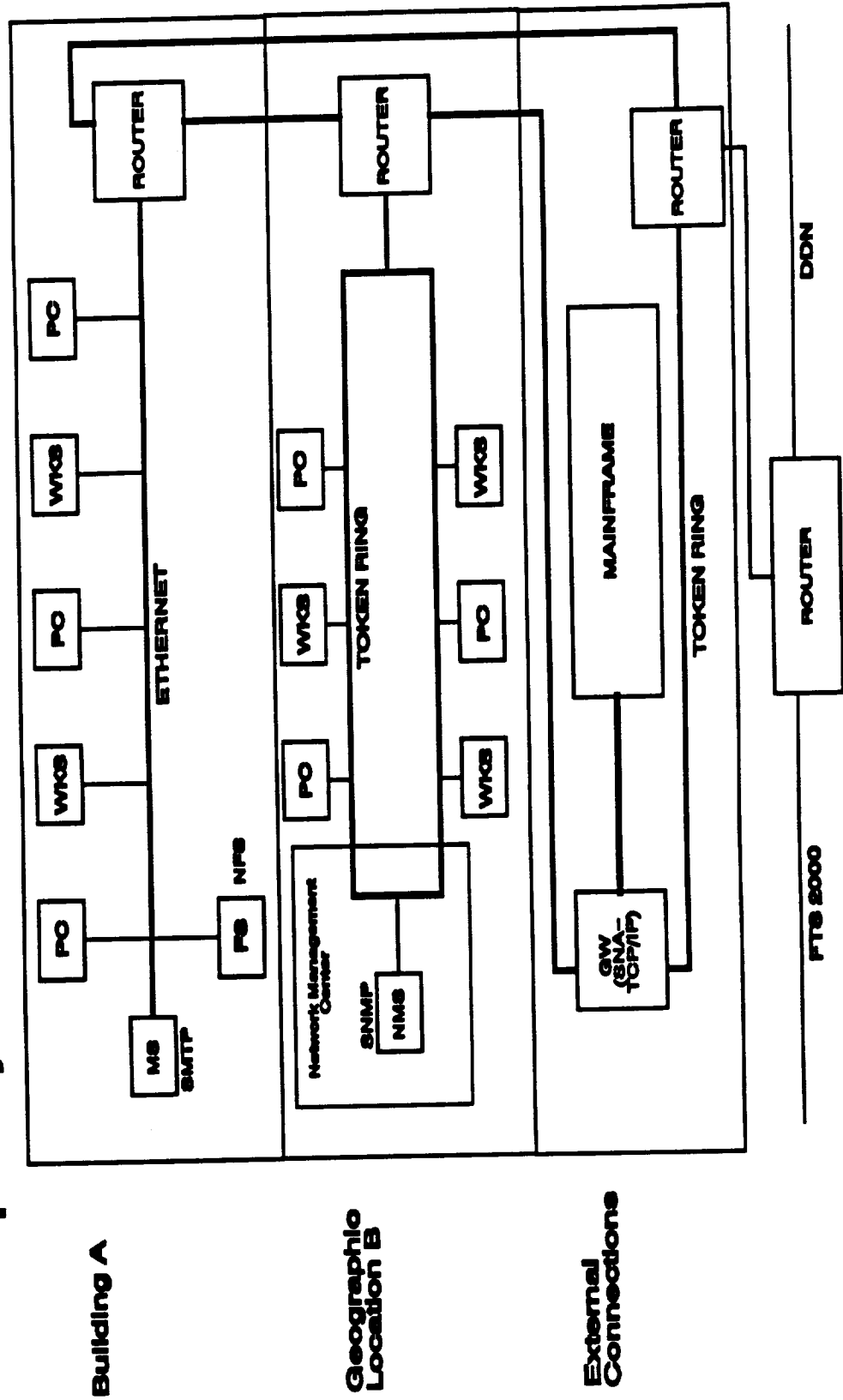
---

## **Transition Strategies**

- **Integrate CASE technology**
- **Provide for centralized data standardization**
- **Promote electronic data and software distribution**
- **Move from batch-oriented to distributed architecture**
- **Buy off-the-shelf rather than build**
- **Use graphical user interfaces**

# OSE PROCUREMENT

## Develop Objective Architecture



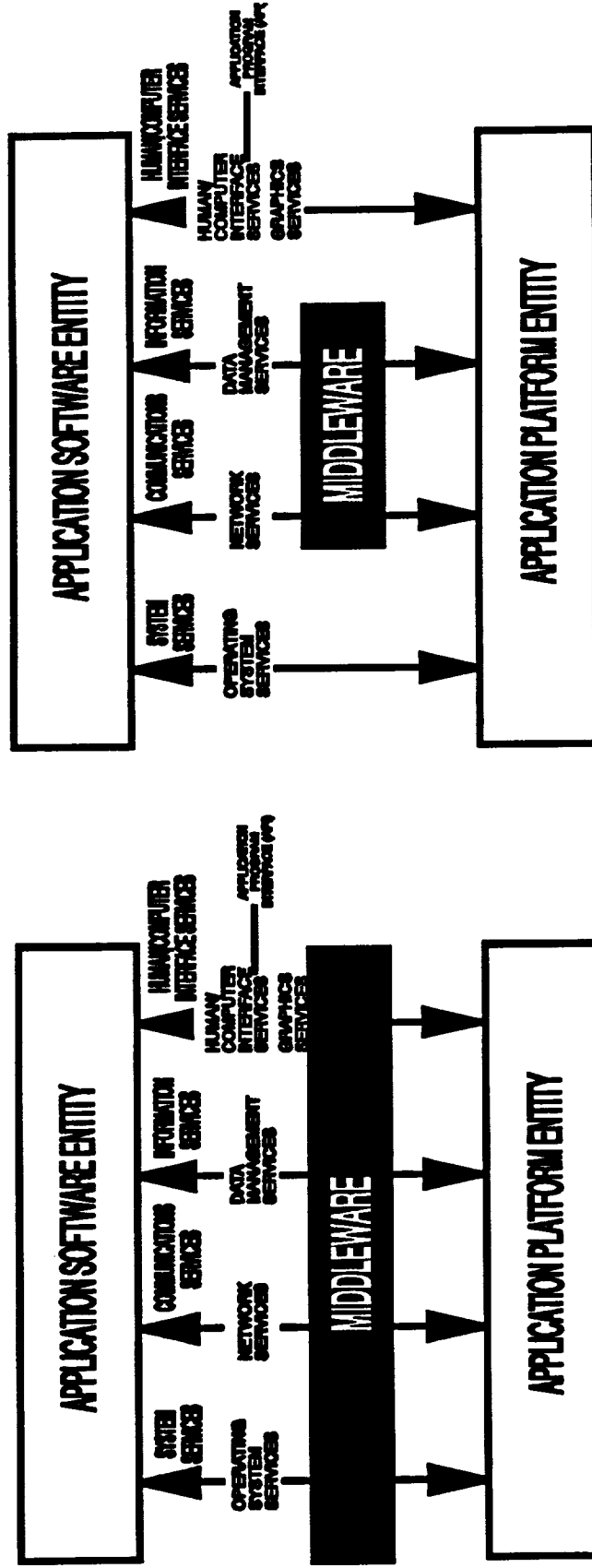
# OSE PROCUREMENT

## Develop Objective Architecture

SERVICE/CONFIGURATION	PERSONAL COMPUTER	WORKSTATION	NETWORK SERVER/ ROUTER/ BRIDGE/ ETC.	DATABASE SERVER	MAINFRAME
Operating System Commands and Utilities System Administration Security					
Human/computer Interface Graphical user Interface Character-based user Interface					
Data Management Database Access Data Dictionary Remote Database Access					
Data Interchange Documents Images Graphics Products					
Programming Languages ISEE 4GL					

# OSE PROCUREMENT

## Transition to Intermediate Targets



# ***OSE PROCUREMENT***

---

## **Annexes**

- **Annex A -- Example RFP Evaluation Evaluation Strategy, SSEB Structure, Example Evaluation Factors/Standards**
- **Annex B -- Example of SOW Requirements Based on Procurement Guide, Office Automation Environment, for illustration only**
- **Annex C -- References (further reading)**

## **Glossary, Index**

# ***OSE PROCUREMENT***

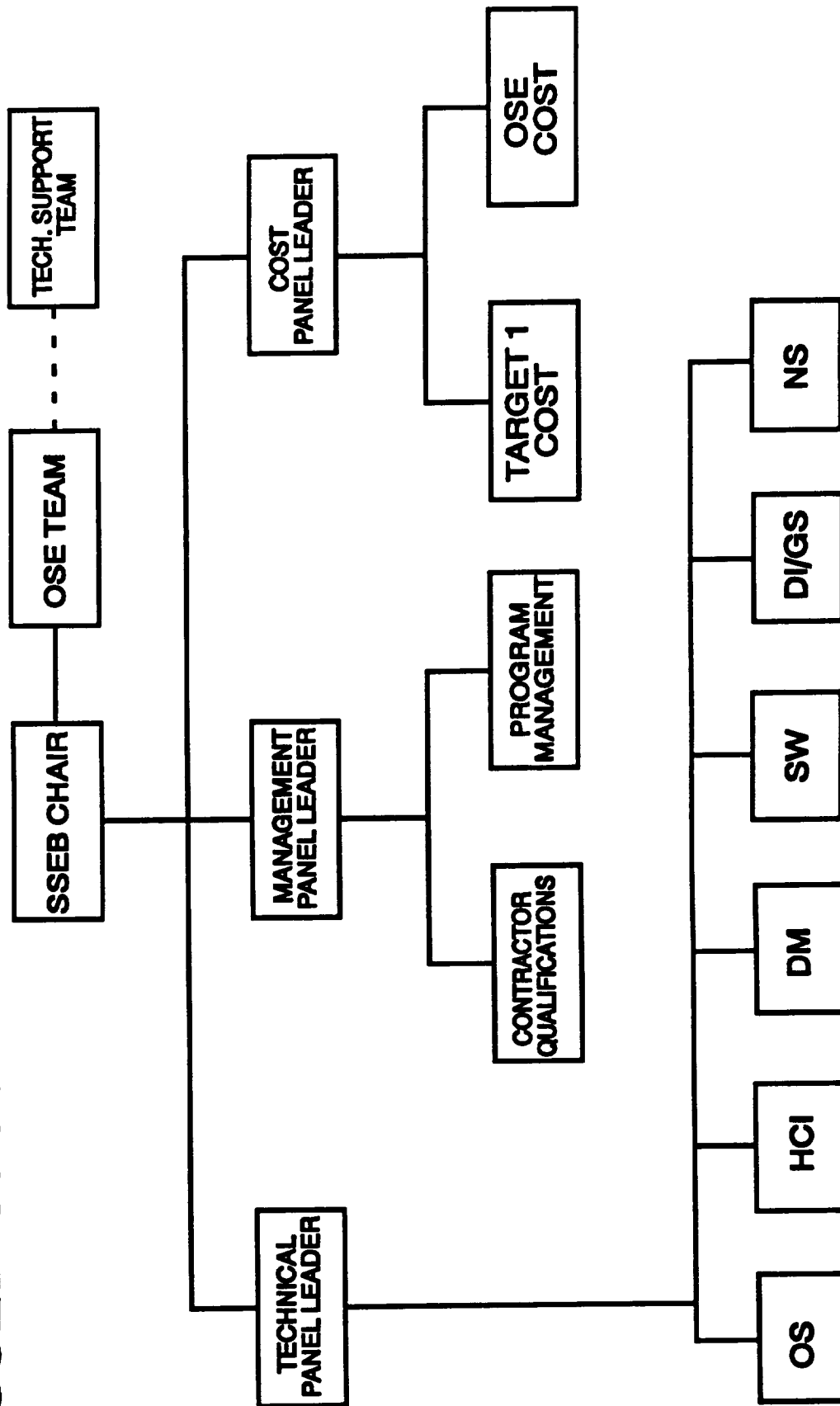
---

## **Evaluation Strategy**

- **Parallel the Application Portability Profile (APP) Version 2.0**
- **Evaluate total management, technology, and cost profile of each contractor**
- **Evaluate transition to the OSE -- Understanding of the environment's complexity, knowledge of OSE and standards, planning and scheduling realism**

# OSE PROCUREMENT

## SSEB Structure



# ***OSE PROCUREMENT***

---

## **Technical Team Evaluators**

### **EXPERTISE REQUIRED**

- System administration
- Operating system internal processing
- Operating system standards
- Software/hardware interfacing
- System security
- Existing organizational operating systems
- Operating system conversion
- Capacity planning and management
- Chargeback systems
- Operations and maintenance
- System auditing

### **POSSIBLE DISCIPLINES**

- System engineer
- Operating system support analyst
- Systems analyst
- System administrator
- Hardware analyst



# ***OSE PROCUREMENT***

---

## **Example Evaluation Factors/Standards**

**Kernel operations are based on FIPS 151-2, and the current Validated Products List (VPL) contains an entry for validation. ....**

**Kernel operations validation was carried out on the proposed platforms. ....**

**Commands and utilities are based on IEEE 1003.2-1993. ....**

**Contractor has demonstrated conformance to the IEEE 1003.2-1993 specification. ....**

**Contractor has demonstrated interoperability with at least one other standard-based product required in this contract.. ....**

## ***OSE PROCUREMENT***

---

### **Example of SOW Requirements**

**[C2] As standards and other specifications required in this contract evolve, the contractor shall provide upgrades for implementations based on the current standards within 12 months of the publication of these standards.**

**[C3] POSIX-like operating system environments for kernel operations offered as a result of this and other requirements in this contract shall implement FIPS PUB 151-2, as a minimum, and shall require validation in accordance with provisions contained in FIPS PUB 151-2.**

# ***OSE PROCUREMENT***

---

## **Summary**

- **Background and Purpose -- PM, SE, lessons-learned, decision model**
- **Organization of Guide -- APP services, standards testing, organizational requirements**
- **OSE requirements -- specifications and applicability, text for SOW**
- **Transition -- plans, baseline, objective, intermediate targets**
- **Annexes -- evaluation strategy/structure, sample SOW, references**

## ***OSE PROCUREMENT***

---

**Agencies must possess in-depth expertise in software development, communications, and database technology in order to evaluate true open systems versus open system marketing. This is still an area where *caveat emptor* applies. An agency can only be prepared.**

# Standard Generalized Markup Language

Robert Donohue

This paper is not available.



# Z39.50 and GILS Model

Eliot Christian

I have some handouts. (The flyer, *Government Information Locator Service (GILS)* and the *January 22, 1994 Draft: Government Information Locator Service (GILS)* are reproduced following the viewgraphs for this presentation.) My presentation is based on that January 22nd draft (Viewgraph 1).

By way of background, where we stand right now is that the Government Information Locator Service was approved yesterday by the Information Infrastructure Task Force. OMB is drafting the bulletin that will give agencies specific direction about when they need to come up on GILS. There are some roles and responsibilities we'll talk about a little bit. The National Institute of Standards and Technologies is establishing the Federal Information Processing Standard, adopting the GILS profile that comes out of the Open Systems Environment Implementor's Workshop, the OIW.

First of all, the objectives (Viewgraph 2). The intention of GILS is that in homes, in workplaces, in schools, in libraries and in hospitals throughout the U.S., the public will be using GILS to discover sources of publicly accessible information maintained throughout the U.S. Federal Government. The agencies will strive to minimize the barriers to "direct users" of GILS. (I'll distinguish in a moment direct users from people who go through intermediary services—it's a very critical piece that's allowing us to get out of the starting blocks here.) There will be a program of evaluation for GILS that will say to what extent it's meeting the service needs of the public, including accessibility, ease of use, accuracy and timeliness of the information, and completeness of coverage. It's not yet clear exactly how that evaluation program will be set up. We are about to get legislation that specifically addresses GILS.

A couple of key concepts here (Viewgraph 3). First of all, GILS is a locator, primarily. That means it is an information resource that identifies other information resources, describes the information available in the referenced resources, and provides assistance in how to obtain the information. GILS encompasses a very wide range of information sources and many different mechanisms for finding and delivering information. It's not a system, except in the sense that the American banking system is a system. It is a set of rules by which we will all play—a set of standards. GILS institutes a collective set of agency based locators. It's not one big centralized thing. It's deliberately decentralized with the belief that you do better if you stay as close as possible to the people who really understand and care for the information, because it's their job, because they're serving their primary user community. So, that's where we base GILS, down in the agencies at the level where people really understand what's in it.

There is, however, a GILS Core that is a mechanism by which the public can view us as a coordinated group of locators, rather than as a bunch of individual fiefdoms. It's basically a

navigation aid to move from locator to locator. So, if you start out at EPA via the Core, you can find out that USGS also has environmental information, things like that. GILS uses network technology— depends on, would not work without, would be an abomination without, network technology. The reason is that that's the only way that we could allow many different views of the information to be on a level playing field. Any other thing you did, you'd have to pre-structure the views so that the user would see something first - something would be on top. With a network, they can see it flat. Another way of putting it is that the user can set the context at the time he's asking his particular question.

Now, many of you deal with the public, and you probably realize most of the public doesn't want to do primary source research. They want pre-digested things. And we anticipate that in GILS. We expect that most of the public need for information will be served through intermediaries— anyone who structures your view ahead of time as opposed to just letting you go out and scarf up the primary sources. Among intermediaries are public libraries and private sector providers, information services like Mead Data Central and BRS, as well as the agencies themselves. Agencies act as intermediaries when they provide a view of the information. If you get earth science information from me, I've decided who I think provides earth science information, and I may have left off the Maharishi Mahesh Yogi, because I don't consider that to be the same kind of science that I do. That's my view—I am an intermediary. I have structured your experience, so you need to be aware of that. GILS makes that distinction very critical.

Another way of looking at it (Viewgraph 4), and this might be old hat to you folks here who are probably sophisticated in this, is that the direct user has an awful lot of flexibility. But the down side is, the direct user has got an awful lot to consider. You've got not only the GILS Core and the other Z39.50 Sources which include huge digital libraries (the Gutenberg project, for example, is going to have a trillion bytes of source material on line). You have things like WAIS and World Wide Web andarchie and Gopher, different views of the world. You have TN 3270. (Does anybody remember—3270 mainframes? A lot of the data is still in the mainframes. Direct users ought to consider that as one of the sources they go after.) You have Virtual Reality. You have video. You have conversations among people. If you're a direct user of research, you have to consider that.

If you're the public who wants some specific answer, you will probably use a product that is structured for that answer. I made up this mythical thing called the Information Master. Somebody selected a certain number of things they think their market wants; they provide that experience through an intermediary service. When they do that, that doesn't require a network. That can be done in print; that can be on fax; that can be done on CD-ROM, on bulletin boards. Direct users, on the other hand, must have the network.

Direct users are assumed to have network access, and to be literate in English to at least a secondary school level. That's because our requirement, as government agencies, is not to translate it into every possible language in every possible schooling level. Intermediaries may do that, but our responsibility is put it out at one particular target audience. We're saying for now, "English at the secondary school level". Direct users must also be capable of using a



personal computer and aware of any limitations of their own hardware or software environment. In building the GILS—putting these rules about how we make our information available to direct users, in effect—what we are doing is building infrastructure (Viewgraph 5).

The Government's role is to set things up so that the diversity of sources can make their information available. This piece of infrastructure is part of the National Information Infrastructure which also includes, you know, moving movies around and deregulation and that kind of stuff. This is one of the government's components. It's something we have a primary role in. It conforms to national, international standards for information and data processing. These two realms, as you are probably aware in the SGI community, have been somewhat divergent in the past. The particular thing we're using here, Z39.50, is kind of a bridge between the two. Although we're adopting an OSI standard, the network services are TCIP, because that's what's out there.

Here are some other design considerations (Viewgraph 6). Particularly as agency people, you might be concerned that GILS will overtake what you're doing. Not at all. It is supplementary. It is not intended to supplant, necessarily, anything you're already doing. Over time, you may well find that GILS serves a need that you were doing some other way and one becomes superfluous, but we see it as something you do in addition to making your information available to your primary user community in the forms you already use. GILS is adopting popular search and retrieval standards, particularly the Z39.50 stuff. That means GILS direct users have access not only to GILS specific stuff, but to other systems such as the NTIS Fedworld system, the GPO Access System, and the National GeoSpatial Data Clearinghouse, which is part of the Spatial Data Infrastructure. Things like the NASA Access Mechanism would be accessible, and some things at Library of Congress, as well as the Global Change Data and Information System. Other government entities at the state, local, foreign, international, non-governmental groups as well, are also picking up on this and making their stuff available the same way.

The functions and contents of GILS are fairly straightforward (Viewgraph 7). First of all, direct users must be able to use non-proprietary standard mechanisms to find the information. By that, we mean software that conforms to the GILS profile at the server side; the profile really just characterizes the behavior of servers. Client software must be able to do ANSI Z39.50 type searches. There will be at least three very large disseminators of free GILS software: NTIS, GPO and the Clearinghouse for Network Information Discovery and Retrieval funded by NSF. Also, Mosaic, by the way, will be able to access GILS sources.

There will be very many other ways that agencies use to organize and present information, things like Gopher or World Wide Web HTTP Servers or TN 3270 or whatever. Those are perfectly legitimate, but these must be in addition to making your GILS records available.

In GILS, we always have information in the locator record about how to order the referenced resource. In some cases, that can be an electronic process; either the ordering or the actual ordering and delivery can be electronic. Of course, in most cases, it's not, because the product

itself may not be electronic or may be large numbers of tapes and that kind of thing. Whatever procedures are defined by the disseminating organization, the description of those procedures is in the GILS locator record.

On the issue of user support services, GILS says nothing about user support services because it's infrastructure. It doesn't ever touch the user. The user should be touched by the agencies who are providing the access to their GILS records or intermediaries other than agencies who are providing access. They have a responsibility to users. The GILS is a set of rules, so we don't actually say anything about user support services. Of course, when we are evaluating later on, one of the things we will be looking at is whether agencies, in their own missions, are doing a good job of supporting people in access to, not only themselves, but to other agencies' stuff, because that's kind of implicit in what you're doing with GILS. You not only put your own stuff up, but you're doing it in common with other agencies, so that's going to be a little bit of a new thing for some agencies.

There will also be a lot of topical directory setup, for example, in bio-diversity or in health care. Different agencies will come together, in many cases with other government or non-government facts, and put common things in a place so that they are searchable for people interested in that topic.

So, the GILS Core is basically defined this way (Viewgraph 8). It's those sources maintained by the U.S. Government. (That's a critical piece—lots of universities are funded by the government and have information, but if it's not maintained by the government, it's not part of the GILS Core) (Viewgraph 9). All of those sources which comply with the Core elements are mutually accessible through interconnected electronic network faculties and do not charge the direct user for the access to the locator information, separate from whether or not there's a charge for access to the referenced resource. This catalog stuff is given away free. You can think of it as advertising. You will be satisfying, by the way, not only your Circular A-130 inventory requirement by doing GILS, but also your electronic records management responsibilities for information systems.

The GILS Core (Viewgraph 10) is estimated to be about 1,000 entries per cabinet or independent agency. So, if NASA gets only 1,000, the Department of Interior gets only 1,000. We have ten bureaus in the Department of Interior; USGS might use only a hundred. At that level, Landsat is an entry; this is highly aggregated stuff. There's lots about Landsat and here's a pointer down into a much more detailed system. So there, again, locator entries are primarily meant as pointers to other information resources where you get a fuller picture of what it is you're interested in (Viewgraph 11). These are the mandatory core elements—mandatory in the sense of when you identify your record as one of those you want to be evaluated against, you better have these fields filled in. The technical profile says how servers must behave to be GILS compliant—it says nothing about the mandatory elements. It's like we have libel laws, but you don't expect your word processor to enforce them. We don't expect the servers to deny access to a record because one of the mandatory fields is empty. The server serves whatever its got. Administratively, we make sure that these things make sense to other processors.

The mandatory elements are kind of obvious stuff from a bibliographic control point of view. Something a little bit different is that this is not a product catalog. It's a catalog of information resources. So, the fact that we have remote sensing images might be identified in a record which you can then see as a subset. It's available at NOAA as this kind of product. It's available at NASA as this kind of product and at USGS as this kind of product. So, GILS gathers the resource into a coherent whole; the products are subfielded. Via a linkage, you would actually be able to hop into the resource that you're talking about. So, when I talk about Landsat and say that we have this thing called the Global Land Information System, I can put a linkage, a URL pointer so that when you click on it in Mosaic, you're into GLIS and, in GLIS you can look for cloud-free images, you can order online, you can do all sorts of stuff like that.

A similar kind of thing is in these optional fields (Viewgraph 12). We have this thing called Cross-Reference. That's a see also kind of pointer. In other words, "I told you about this particular resource and I may have actually let you go down into it. Here are some other related things you might want to know about." In other words, if you are looking at Landsat stuff, you might also want to look at the NEXRAD stuff that we have.

These other things here are fairly straightforward. Agency supplemental is the place where agencies add in whole bunches of other stuff that they couldn't find any other place to fit. Although I've described some elements here, agencies may add any other locally defined tags anywhere in the structure at any time they want. So, record by record, you could say, for this particular one, "I'd like to add the acronym, for this one over here. I have a field called *data category* that I'd like to report." In fact, that's what I'd like to show you now. Kind of what the records look like.

First, let me just put up sort of a conceptual thing (Viewgraph 13). This is where we're going, what we're trying to achieve—seamless access so that people don't constantly trip over the differences between agencies, the differences between access mechanisms, all kinds of differences. Typically, I would see you starting out with an agency that you know. Via that agency, you may find the pointer to the GILS Core where the other agencies have referenced similar things in common with you. For example, I might start out with the Earth Science Data Directory. In that, you would find a pointer off to the GILS Core that would help you find other federal agencies who do similar kinds of things. Linkages there could take you to the electronic visitor center, Mosaic Home Pages, where you give your view of the agency as though people had just walked in the door. You've seen those before, I'm sure. Other topical directories, depending on what your searching in—there might be lots of those. You can walk over to product inventories, which might be electronic delivery, or they might be a mail-in-your-order kind of thing. Because we're using standards used at digital libraries, you can walk off into contents of things like at the Library of Congress or some online exhibits which are kind of neat. You have all these different things out there. GILS, because it's adopting a common set of rules, makes much of that commonly accessible for GILS direct users, if you have good client software. Now remember, of course, you could also see GILS in a printed form, in which case you see the information, but you can't click on the page and do anything. You can take that information, write it on a napkin and go somewhere else.

I would like to show you what a GILS record looks like if you bring it up, for example, in Mosaic. This is one that we've created. We got a test data set out there with 26 USGS records. We've indexed them using WAIS. So, WAIS responds to Z39.50 requests. So, from that point of view, you can do a search and get hits. This particular one is the USGS server for Gopher. You see a title and an originator. Here's a locally defined field that I introduced—*acronym*. (The actual way I present these records or I explored these records for my database, by the way, is in SGML, which is a popular way to pass this stuff around, because the structuring is nice and neat.) In this particular case, I'm putting out an HTML record, so I have the ability to express hypertext links via this anchor. Here I have a link to the actual Gopher server that I am describing. That is a linkage, because you are going down into the thing described. I also have, in this particular record, a cross-reference. In here, the cross-reference is to an HTTP server for World Wide Web. So, here you have a cross-reference over to my Mosaic Home Page.

I can give you another example. This is an actual information resource—*Aerial Color Photographs of Metropolitan Areas*. Here we have the spatial reference giving the bounding rectangle around it so that people can do a search spatially. In Mosaic, spatial search is not there right at the moment. In many other clients, it is. People are going to get different functionality, depending on the clients they get. Ultimately, clients should disappear and become part of your application. If you're doing GIS work, the GIS should go out and find things for you. When you're doing a File Find on your hard disk, it should consider the whole world, not just your own hard disk. We're going to be getting there, but it's going to be a couple of years. So right now, you use these things called clients. The fact is that it's not really imbedded in your day-to-day work yet. But it will be. I can show you what the SGML that generates this looks like if you're interested. It's actually pretty straightforward. You can actually maintain these records in whatever you want. In many cases, agencies will already have this stuff. It's the same kind of stuff you do in your budget briefing books. You tell Congress what you're doing. If you have money and you're still doing it, you must have told them at some point. It's the kind of things you have been giving to NARA, for your electronic records management. And of course, you could generate these using things like sophisticated database systems or you could use D-Base. You can just as easily use word processing; you know, a thousand entries is not a lot of typing, particularly when they're only a thousand words each.

So, here is some SGML. I happened to have written this in Microsoft Access and wrote an exporter that exports them in HTML, SGML, and, SUTRS. (SUTRS is the other format that is required to be supported by GILS servers: Simple, Unstructured Text Records Syntax.) This is what SGML looks like for these.

We start each record off with a <REC> and then this is a field title and then you close it with a slash title. We should have a formal DTD for this, but right now, it's just implicit. In the example done here with "abstract," I've opened another field before closing it. That is how you represent that format is a subfield or it's nested within abstract. You have to close format before you close abstract.

HTML is the same sort of stuff, except it doesn't preserve this naming of the tags. It uses an implicit tagging definition that just gives you certain functions. SGML is the superset. It's the more powerful way to actually represent structure unambiguously and reversibly, so that I can pull these things back and still know that that must have been the format field, because it still says "format" in it. When I put them out in HTML, it says, "This is going to be a descriptive list." The fact that it actually was this particular field is now lost, because it's over in text somewhere.

WAIS, doesn't understand fields. WAIS treats everything like a big blob conglomerated with lots of other blobs. When you're looking for things, you get a sense of where things are statistically. WAIS looks real nice—you get things back. It's not the same as having an understanding of the semantics that went with this stuff. It doesn't make any attempt to preserve that. And from the point of view of GILS, that's fine. With this kind of record, you'll get good hits. You'll get the kind of things you're looking for. WAIS is not the only solution—it is among the range of solutions.

Let's take a quick look at what a SUTRS record looks like. SUTRS is just simple unstructured text. The rules for constructing SUTRS, that are in the GILS profile, say you will always give the actual name of the field, a colon, a space and then the content. You'll have carriage return line feeds and lines that won't be more than 80 characters across. Because we offer up SUTRS, dumb clients can simply grab that information and just display it to the user and not have to have any understanding.

Two other things that GILS profile requires: one, you have to be able to serve MARC records for this stuff, and the other is you have to be able to serve up what is called "Generic Record Syntax." Generic Record Syntax simply means that, on request, I will give to you as much as I knew about the record. So, we won't have a loss of information as intermediaries copy from each other and move the information out down the chain. In other words, you can get everything I knew about it so you could reproduce it. It's not true, however, that that second intermediary can reproduce the MARC record, because the transform to MARC loses information. All that we knew about the record as we had it in the construct is what was actually in the original server.

# **Toward a Government Information Locator Service (GILS)**

Presented by Eliot Christian

U.S. Geological Survey

*(based on GILS draft 1/22/94)*



# **GILS Objectives**

- **Public uses GILS to discover sources of information throughout Federal government**
- **Includes publicly accessible information**
- **Ensures reasonable ability to access the information**
- **Minimizes barriers to direct users (e.g., equipment, software required, cost, technical complexity)**
- **Protects rights of privacy and intellectual property**
- **Evaluates how well objectives are being met**



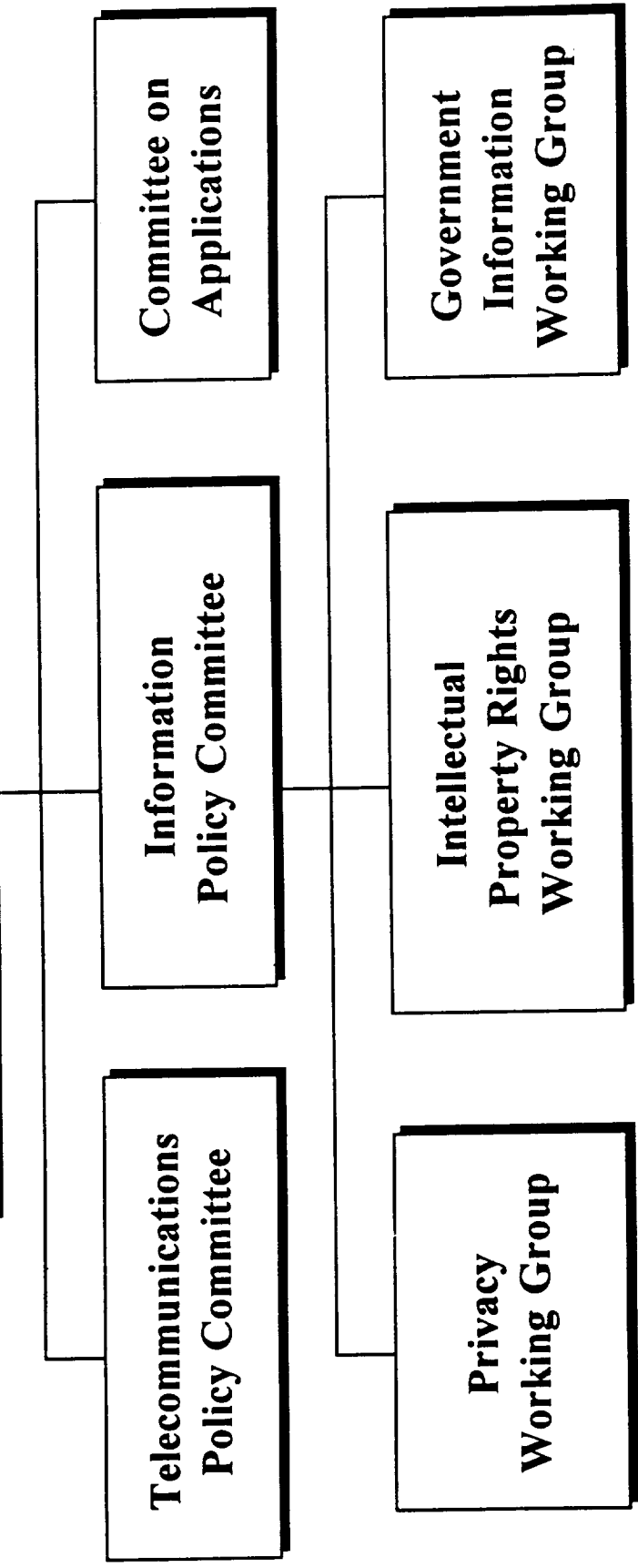
## Background

- 1992 OMB/NARA/GSA study (Charles McClure, Syracuse University) recommends Internet, Z39.50
- Administration's strategic technology policy "Technology for America's Growth"
- OMB Circular A-130 "Management of Federal Information Resources"
- A-130 Next Steps:
  - Improve electronic mail among Federal agencies;
  - Convert paper forms to electronic access;
  - Promote establishment of agency-based GILS





# **Information Infrastructure Task Force**



## Key Concepts

- Identifies other information resources, describes information available, and how to obtain it
- Builds on agency-based locators (i.e., defines the common approach with decentralized operations)
- Uses network technology, but GILS contents available via e-mail, bulletin boards, and off-line media such as CD-ROM and printed publications
- Public uses GILS directly or through intermediaries such as agencies, public libraries, private sector



# Direct Users

GILS Core  
All GILS  
All 239.50 Sources  
WAIS  
World Wide Web  
Gopher/ Veronica  
Archie  
FTP  
Telnet  
TN 3270 Data Systems  
PH  
E-mail  
News  
Libraries  
Discussion Groups  
M Bone  
C U See Me  
Virtual Reality

Welcome to the  
Information Master.

You can search by  
one of the following:

Agency  
Subject  
Originator  
Location  
Date  
Event  
Personal Name

**Direct users have flexibility,  
but much to consider**

**Intermediaries provide  
a more focused experience**

# **National Information Infrastructure**

- **GILS is one component of the U.S. National Information Infrastructure**
- **Conforms to national and international standards for information and data processing**
- **Network services focus on the Internet initially**
- **Expected to enhance technical innovation for public access**



## Other Design Considerations

- Supplements other agency and commercial information dissemination mechanisms
- Other Federal and non-Federal information sources will be accessible through GILS
- Other government entities encouraged to institute compatible locators
- Will accommodate the expressed needs of other government entities, where practical



## **Functions and Content**

- Direct users of GILS must be able to use non-proprietary, standard mechanisms (i.e., ANSI Z39.50 compatible software)
- Always provides ordering information, can also support electronic delivery
- User support services are not specifically prescribed within GILS
- Includes some redundancy; Users may get different perspectives on the same question
- Federal agencies develop subject-oriented directories that supplement GILS Core



# Core Requirements

- Those sources maintained by the U.S. government, all of which:
  - comply with the defined Core Elements,
  - are mutually accessible through interconnected electronic network facilities
  - do not charge direct users
- Support NARA and Circular A-130 inventories
- Designed to minimize conflict with USMARC



# **GILS Core**

**Services for finding government information take many forms**

**Many kinds of finding aids for information**

**Locators that are electronic**

**Digital locators**

**Network-based locators**

**Internet-based locators**

**Z39.50 information servers**

**GILS profile servers**

**GILS Core**



# **GILS Core Contents**

- **References each Federal information system holding publicly accessible data or information**
- **Lists at highly aggregated level (e.g., 1,000 entries per Agency, entries under 1,000 words)**
- **Defines content standards for Core Elements**
- **Defines GILS Core object transfer standard, compliant with ANSI Z39.50**
- **Provides preferred display formats for print or electronic presentation**



# Mandatory Core Elements

- Title
- Control Identifier
- Abstract
- Purpose
- Originator
- Access Constraints
- Use Constraints
- Availability
  - Distributor, Resource Description, Order Process, Technical Prerequisites, Time Period, Linkage, Linkage Type
- Point of Contact
- Record Source
- Date Last Modified

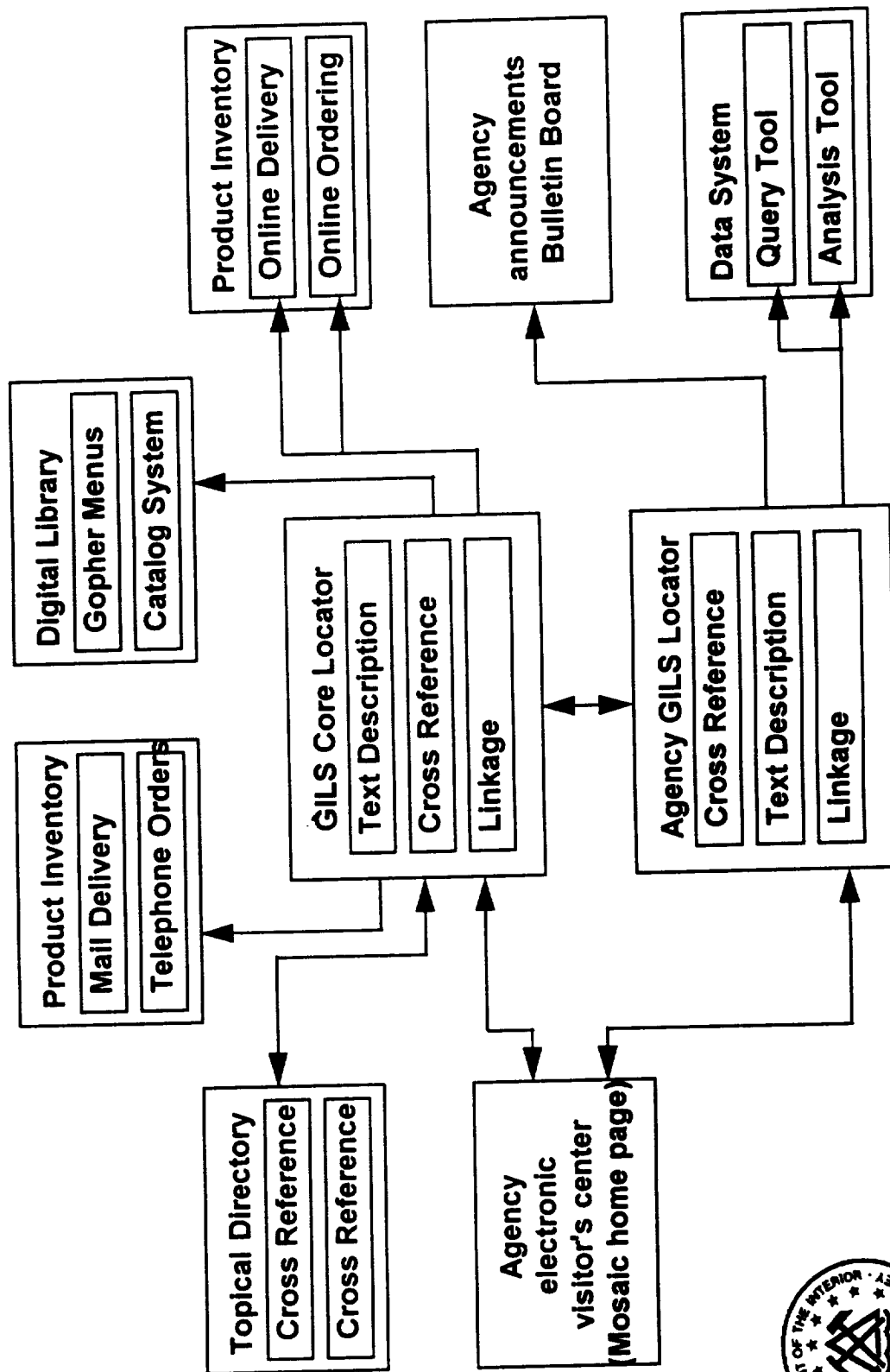


# Other Core Elements

- Required for Information Systems
  - Agency Program
  - Sources of Data
- Optional
  - Controlled Vocabulary
  - Local Subject Index
  - Methodology
  - Spatial Reference
  - Time Period of Content
  - Cross Reference
  - Original Control Identifier
  - Agency Supplemental



# Seamless Access



## Government Information Locator Service (GILS)

As part of the National Information Infrastructure, the U.S. Federal government is proposing a Government Information Locator Service (GILS) to help the public locate and access information. An Office of Management and Budget Bulletin will be published this year to provide implementing guidance specifying Federal agency responsibilities. The National Institute of Standards and Technology will also establish a Federal Information Processing Standard specifying a GILS Profile with mandatory application for Federal agencies establishing locators for information.

### What is GILS?

"Every year, the Federal Government spends billions of dollars collecting and processing information (e.g., economic data, environmental data, and technical information). Unfortunately, while much of this information is very valuable, many potential users either do not know that it exists or do not know how to access it. We are committed to using new computer and networking technology to make this information more accessible to the taxpayers who paid for it. In addition, it will require consistent Federal information policies designed to ensure that Federal information is made available at a fair price to as many users as possible while encouraging growth of the information industry."

*"Technology for America's Economic Growth,  
A New Direction to Build Economic Strength"*

GILS would identify public information resources throughout the Federal Government, describe the information available in those resources, and provide assistance in obtaining the information. It would consist of a decentralized collection of agency-based information locators and associated information services. GILS would supplement, but not necessarily supplant, other agency information dissemination mechanisms and commercial information sources.

The public would be served by GILS through intermediaries or directly. Central disseminating agencies such as the Government Printing Office and the National Technical Information Service would act as intermediaries to GILS, as would Depository Libraries, other public libraries and private sector information services. Access to GILS contents could also be accomplished through kiosks, "800 numbers," electronic mail, bulletin boards, FAX, and off-line media such as floppy disks, CD-ROM, and printed works.

While GILS would encompass a very wide range of information sources and many mechanisms for finding and delivering information, a "GILS Core" would be specifically defined to be a definitive locator of agency information resources. The GILS Core would be accessible on public networks without charge to direct users.

GILS would use network technology and the American National Standards Institute Z39.50 standard for information search and retrieval so that information can be retrieved in a variety of ways, and so that GILS direct users can ultimately gain access to many other major Federal and non-Federal information resources. GILS would also include automated linkages that facilitate electronic delivery of off-the-shelf information products, as well as guide users to data systems that support analysis and synthesis of information.

## **OMB Circular A-130 and Information Locators**

On June 25, 1993, the Office of Management and Budget revised Circular A-130, "Management of Federal Information Resources," to strengthen policies for managing government information (58 F.R. 36068, July 2, 1993). Circular A-130 states that availability of government information in diverse media, including electronic formats, permits the public greater flexibility in using the information, and that modern information technology presents opportunities to improve the management of government programs to provide better service to the public. It notes that the development of public electronic information networks, such as the Internet, provides an additional way for agencies to increase the diversity of information sources available to the public, and that emerging standards such as ANSI (American National Standards Institute) Z39.50 will be used increasingly to facilitate dissemination of government information in a networked environment.

Circular A-130 states that agencies shall:

- Disseminate information products on equitable and timely terms;
- Avoid establishing exclusive, restricted, or other distribution arrangements that interfere with the availability of information dissemination products on a timely and equitable basis;
- Use voluntary standards and Federal Information Processing Standards;
- Use electronic media and formats, including public networks, as appropriate and within budgetary constraints, in order to make government information more easily accessible and useful to the public;
- Take advantage of all dissemination channels, Federal and nonfederal, including State and local governments, libraries and private sector entities;
- Provide information describing how the public may gain access to agency information resources;
- Help the public locate government information maintained by or for the agency;
- Establish and maintain inventories of all agency information dissemination products;
- Develop such other aids to locating agency information dissemination products including catalogs and directories...

### **Where to find more information on the Government Information Locator Service (GILS)**

Based on the work of interagency groups such as the Working Group on Public Access, and in coordination with the Information Infrastructure Task Force (IITF), the Office of Management and Budget has endorsed a vision document describing how GILS may be implemented. The document will become a report to the IITF after review by the three IITF Committees and the United States Advisory Council on the National Information Infrastructure. Prior versions of the document were reviewed by various Federal agencies and other interested parties, including some non-Federal organizations and by the general public through notices in both the Federal Register and the Commerce Business Daily, as well as through a public meeting held at the Department of the Interior on December 13, 1993.

The GILS document is available on the FedWorld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at 130.11.48.107 as /pub/gils.doc (Microsoft Word for Windows format), /pub/gils.wp (WordPerfect 5.2 format), or /pub/gils.txt (ASCII text format). Comments should be sent by electronic mail to [echristi@usgs.gov](mailto:echristi@usgs.gov), or on paper to Eliot Christian, U.S. Geological Survey, 802 National Center, Reston, VA, 22092.

## *January 22, 1994 Draft*

# **Government Information Locator Service (GILS)**

---

The Office of Management and Budget (OMB), in coordination with the Information Infrastructure Task Force (IITF), is promoting the establishment of an agency-based Government Information Locator Service (GILS) to help the public locate and access information throughout the U.S. government.

This document presents a vision of how GILS may be implemented. It is intended to be issued as a report to the IITF after review by the IITF Committee on Information Policy, the IITF Committee on Telecommunications Policy, the IITF Committee on Applications and Technology, and the United States Advisory Council on the National Information Infrastructure.

This document was developed primarily by Eliot Christian and the Locator Subgroup of the Interagency Working Group on Public Access. Prior versions of this document were reviewed by various Federal agencies and other interested parties, including some non-Federal organizations and by the general public through notices in both the Federal Register and the Commerce Business Daily, as well as through a public meeting held at the Department of the Interior on December 13, 1993.

The design of GILS follows generally on the work of Dr. Charles McClure of Syracuse University as described in the 1992 report to OMB, the National Archives and Records Administration, and the General Services Administration, entitled "Identifying and Describing Federal Information Inventory/Locator Systems: Design for Network-Based Locators."

This document is available on the FedWorld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at 130.11.48.107 as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).

Comments should be sent by electronic mail to [echristi@usgs.gov](mailto:echristi@usgs.gov), or on paper to Eliot Christian, U.S. Geological Survey, 802 National Center, Reston, VA, 22092.





# **Government Information Locator Service (GILS)**

**DRAFT**

**Report to the Information Infrastructure  
Task Force**

---

## Table of Contents

1. Context .....	1
2. GILS Overview	
2.1 Characteristics of GILS .....	3
2.2. GILS from the User Perspective .....	4
2.3. GILS from the Provider Perspective .....	5
3. Service Requirements	
3.1 Design Principles .....	6
3.2 Functions .....	7
4. Core Requirements	
4.1 Functions .....	8
4.2 Content .....	9
4.3 Core Element Definitions .....	10
Appendix 1: Glossary .....	1-1
Appendix 2: Extracts from Draft GILS Profile .....	2-1

---

### **1. Context**

The Administration's Strategic Technology policy document entitled "Technology for America's Economic Growth, A New Direction to Build Economic Strength" states:

Every year, the Federal Government spends billions of dollars collecting and processing information (e.g., economic data, environmental data, and technical information). Unfortunately, while much of this information is very valuable, many potential users either do not know that it exists or do not know how to access it. We are committed to using new computer and networking technology to make this information more accessible to the taxpayers who paid for it. In addition, it will require consistent Federal information policies designed to ensure that Federal information is made available at a fair price to as many users as possible while encouraging growth of the information industry.

On June 25, 1993, the Office of Management and Budget (OMB) revised Circular A-130, "Management of Federal Information Resources," to strengthen policies for managing government information (58 F.R. 36068, July 2, 1993). Circular A-130 encourages agencies to use new technologies to make government information available to the public in a timely and

equitable manner via a diverse array of sources, both public and private. It states that availability of government information in diverse media, including electronic formats, permits the public greater flexibility in using the information, and that modern information technology presents opportunities to improve the management of government programs to provide better service to the public. It also notes that the development of public electronic information networks, such as the Internet, provides an additional way for agencies to increase the diversity of information sources available to the public, and that emerging standards such as ANSI (American National Standards Institute) Z39.50 will be used increasingly to facilitate dissemination of government information in a networked environment.

OMB Circular A-130 states that agencies shall:

- Disseminate information products on equitable and timely terms;
- Avoid establishing, or permitting others to establish on their behalf, exclusive, restricted, or other distribution arrangements that interfere with the availability of information dissemination products on a timely and equitable basis;
- Use voluntary standards and Federal Information Processing Standards where appropriate or required;
- Use electronic media and formats, including public networks, as appropriate and within budgetary constraints, in order to make government information more easily accessible and useful to the public;
- Take advantage of all dissemination channels, Federal and nonfederal, including State and local governments, libraries and private sector entities;
- Provide information describing how the public may gain access to agency information resources;
- Help the public locate government information maintained by or for the agency;
- Establish and maintain inventories of all agency information dissemination products;
- Develop such other aids to locating agency information dissemination products including catalogs and directories...

In addition to the Strategic Technology policy and the strengthened Federal policy concerning information dissemination, the Administration has called for a more active role of agencies in strengthening the implementation of the Freedom of Information Act (FOIA). The belief is, if agencies actively open up access to information, the use of formal FOIA requests by the public will become less necessary thereby improving agency responsiveness and decreasing costs.

The responsibilities of Federal agencies with regard to the management of electronic records are also growing in importance as their reliance on electronic information systems increases. The National Archives and Records Administration (NARA) will be issuing revised guidance to agencies to update policies consistent with 44 U.S.C.

Because it is essential to the operation of government and to democratic principles that agencies actively manage information, these and other laws and policies assert a fundamental requirement that Federal agencies maintain readily accessible inventories of their records and other information holdings. To help the public locate and access public information within agency inventories, the Administration has committed to promote the establishment of an agency-based Government Information Locator Service (GILS).

Agencies are already required to create and maintain an inventory of their information systems and information dissemination products under 44 U.S.C., FOIA, and OMB Circular A-130. Although compliance with these requirements varies greatly, the incremental cost of making those inventories accessible through GILS is expected to be minimal. Accordingly, the participation of agencies in establishing and maintaining the GILS Core may be accomplished as a collective effort executed within existing funds and authorities.

OMB expects to publish in 1994 an OMB Bulletin that would follow-on Circular A-130 and provide implementing guidance specifying agency responsibilities to participate in GILS and setting performance measures. The National Institute of Standards and Technology (NIST) will establish a GILS Profile as a Federal Information Processing Standard (FIPS) with mandatory application for Federal agencies establishing locators for government information. A program of evaluation will be established to evaluate the degree to which GILS meets user information needs, including factors such as accessibility, ease of use, suitability of descriptive language, accuracy, consistency, timeliness, and completeness of coverage.

## **2. GILS Overview**

### **2.1 Characteristics of GILS**

In homes, workplaces, schools, hospitals, and libraries throughout the United States, the public should be able to discover sources of publicly accessible information maintained throughout the U.S. Federal government. To meet that goal, Federal agencies are organizing the agency-based GILS as a component of the National Information Infrastructure (NII).

GILS must be many things to many people. It must be comprehensive, yet user friendly. It must answer specific questions, yet allow for scanning a wide range of government information. It must be able to answer questions from the most naive users, yet allow for in-depth research as well. It also must be of direct service to the public, yet not undermine the diversity of existing information sources. GILS must reflect an inclusive policy that lets any private sector information provider which is providing GILS sources to make its own resources known and accessible.

GILS depends critically on other aspects of the emerging NII. GILS must be implemented with full recognition of individual privacy and intellectual property rights. Agencies will need to ensure that members of the public whom the agency has a responsibility to inform have a reasonable

ability to access GILS and the underlying information resources and information dissemination products. Agencies participating in GILS must take care to minimize barriers to use, including equipment and software requirements, cost, and technical complexity.

## **2.2. GILS from the User Perspective**

The public will use GILS either directly or through intermediaries. In an exploration analogy, the distinction is that direct users roam at will but users of intermediate services take a guided tour. The following are some examples of GILS direct users and intermediaries:

- A researcher interested in national health care may access a wide range of GILS sources as a direct user in order to explore issues from virtually any perspective.
- An educator interested in keeping up with electronic educational materials may access a few GILS sources once a month as a direct user over a dial-up connection to the Internet.
- An information service may access GILS hourly as a direct user, and also act as an intermediary by constructing a value-added directory derived from GILS for sale to users with specific needs such as economic forecasts.
- A network service provider may offer an intermediate service by offering selected GILS access to users as a set of options within their bulletin board services.
- A Federal agency may act as an intermediary in adding GILS access into its existing information service to provide public information referrals to sources in other agencies.

A major advantage of the networked and decentralized design of GILS is that it allows direct users to explore many different perspectives of government information. Since they are less constrained in their searching, direct users have more flexibility to explore the full complement of available information. However, direct users must have network access and they are also assumed to be literate in English to at least the secondary school level, capable of using a personal computer, and aware of any constraints of their own hardware or software environment.

In contrast, intermediate services are typically oriented toward a particular user community and present a more focused experience to users searching for information. Intermediate services need not require users to have network access, but can present GILS information in the full range of communications media. Such services can be offered via electronic mail, bulletin boards, FAX, and other media such as CD-ROM (Compact Disk-Read Only Memory), printed publications, telephone help desks, and information kiosks in public places such as envisioned in the Administration's Service to the Citizen initiative.

Clearly, most of the public need for access to government information will be well served through the diverse array of public and private sector service providers. Casual users and those lacking network access will be served typically through products and services offered by agency or non-government intermediaries such as Depository Libraries, other public libraries, and private sector providers. These intermediaries obtain GILS information either as direct users themselves

or from other intermediaries, but the extent of government information that may be provided by any particular intermediate service is not prescribed by GILS.

### **2.3. GILS from the Provider Perspective**

The design of GILS follows generally the work of Dr. Charles McClure of Syracuse University as described in a 1992 report to OMB, NARA, and the General Services Administration (GSA). A locator is here defined as an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in how to obtain the information.

A key concept of GILS is that it uses network technology to support many different views across many separate locators. Although directly accessible on networks, all or part of the GILS contents can also be made available by intermediaries through other media. These alternative mechanisms help assure that the information is available through a diversity of sources, both public and private, and covering the full range from telephone help through print media and up to the most sophisticated electronic network technologies.

GILS organizes a collective set of agency-based locators and associated information services that are decentralized so that responsibilities stay as close as possible to those who understand and care for the information and who are serving the agency's primary user community. Each agency is responsible for ensuring that its GILS components are continuously accessible to GILS direct users, whether through agency computer resources or through other arrangements. Certain agencies also have in their primary mission an additional role in helping the public to access information maintained elsewhere in the government.

Among the GILS agency components is a set of locator records designated to comprise the GILS Core. The GILS Core consists of those locator records that are required to be maintained by those Federal agencies having significant information holdings, each of which describes agency holdings. These agency locator records can be aggregated by direct users of GILS to provide a broad view of all Federal government holdings, and they can also be combined in other ways because GILS uses interoperable standards for information search and retrieval.

Agencies such as the Government Printing Office (GPO) and the National Technical Information Service (NTIS), as well as private sector information providers, can supplement access to the GILS Core with access to other Federal and non-Federal information. Other major Federal government information systems such as the GPO Access System, the NTIS FedWorld system, the National Geospatial Data System, and the Global Change Data and Information System will be accessible to GILS direct users. GILS direct users may have access to a wide range of additional Federal information on the network such as current and historical information on Federal programs and institutions; public notices; law, regulation, policy, and procedural materials; and listings of experts and office locations. Other government entities (State, local,

foreign, international) and non-government organizations will also be encouraged to institute locators compatible with the international standards used in GILS. GILS itself will accommodate the expressed needs of other government organizations where practical.

### **3. Service Requirements**

#### **3.1 Design Principles**

GILS is a component of the NII that is evolving with guidance from the Information Infrastructure Task Force. GILS will be interoperable with other component NII initiatives such as the National Spatial Data Infrastructure. GILS is also expected to adapt to and to encourage technical innovation, especially in ways that enhance public access.

GILS will conform to national and international standards for information and data processing. Participants in GILS will use voluntary standards processes (e.g., ANSI, the Open Systems Environment Implementors Workshop, and the Internet Engineering Task Force) to promote interoperability of search and retrieval mechanisms, network communications, user authentication, and resource identifiers, among other essential components. Near-term implementations of GILS will use the Internet and its communications protocols, but GILS is based on the international Open Systems Interconnection (OSI) model in order to be compatible with a wide range of technologies. The application profile specifying GILS compliance will be maintained and published by NIST.

GILS takes advantage of the network technology known as client-server architecture, which allows information to be distributed among multiple independent information servers. Client applications may allow the user to question many servers concurrently and have the answers automatically combined. In this way, GILS allows for agencies to maintain various information resources optimized for their usual customers, yet the resources can be rapidly collated in a different way to serve a different need. Special provisions are made in GILS to support navigation among GILS locators by using hierarchical browsing as well as textual searching.

GILS supports seamless access not only among locators but directly to referenced information resources. When implemented at both the client and server, GILS linkages facilitate electronic delivery of off-the-shelf information products, as well as connection to data systems that support analysis and synthesis of information.

GILS does not directly address the general problem of how to correlate or otherwise combine data gathered from among sources that are maintained separately. Communities of interest, such as the participants in the National Spatial Data Infrastructure, are working toward improving the situation, but no general solution has yet met with wide acceptance. While there are deep and complex issues surrounding data comparability, it is clear that complete and readily accessible

documentation will be a key requirement. GILS does provide a basis for broad accessibility to the highest level documentation of data and information holdings.

### **3.2 Functions**

Because GILS builds on agency-based locators, supplementing other agency and commercial information dissemination mechanisms, user support services are not specifically prescribed. Federal agencies are required to provide an appropriate level of user support services for their components of GILS, either directly or through intermediaries.

Requests and arrangements for delivery of information located through GILS are handled in a variety of ways, including support for electronic delivery of information products. Much of the referenced information is not available in electronic form, although the trend is clearly in the direction of electronic network availability. At a minimum, GILS always provides information regarding request and delivery procedures for the various distribution options as defined by the disseminating organization.

Direct users of GILS must be able to use non-proprietary, standard mechanisms to discover information sources and retrieve basic textual information content. This function is within the scope of the information search and retrieval standard known in the United States as ANSI Z39.50 and internationally as ISO (International Standards Organization) 10162/10163.

GILS locators must be accessible on interconnected electronic network facilities and must support the currently approved ANSI Z39.50 standard for information search and retrieval. To facilitate interoperability of independently developed components of GILS, such as discrete client and server software, a GILS Profile is being drafted by a research project between the U.S. Geological Survey and Syracuse University, funded by the Interagency Working Group on Data Management for Global Change. (Extracts from a recent draft of that specification are included as Appendix 2 to this document.) This research effort is intended to lead to a formally approved GILS Profile.

The GILS Profile will provide a complete specification of GILS as it makes use of ANSI Z39.50, but also specify where necessary those characteristics of GILS that are not within the scope of ANSI Z39.50. The GILS Profile will provide for navigating among GILS locators through the specifications given for the GILS Core Elements. The GILS Profile will not constrain how information is maintained at the source, nor how the information is displayed to the user.

Access to GILS is expected to be embedded within many different computer applications, ranging from the very simple to those that support conceptual search across languages, dynamically interpret natural language, or filter search requests to sift huge amounts of information automatically. Software conforming with ANSI Z39.50 must also conform to the GILS Profile to provide full functionality to GILS direct users. Public domain client software that supports access



to GILS will be available from GPO, NTIS, and the Clearinghouse for Networked Information Discovery and Retrieval, among others.

Alternative ways to organize and present networked information are encouraged, but agencies participating in GILS will implement such alternatives in addition to supporting access by GILS direct users who employ the currently approved ANSI Z39.50 standard. For example, information organized via the OSI X.500 standard can be made accessible via ANSI Z39.50, thereby enhancing access capabilities. It should also be noted that GILS direct users will typically have access to a wide variety of information sources that do not comply with the GILS profile but which are compliant with various other standards.

Some internal redundancy in GILS is to be expected. Such redundancy is appropriate because the same information resources may be described differently to different audiences or for different purposes, and descriptions will cover information resources at a wide range of aggregation. Also, the same information resources may be described differently by different information services that participate directly or as intermediaries in providing Federal information to the public. Because GILS incorporates a variety of automated and manual search techniques, users will obtain different perspectives on a question depending on how GILS is used.

GPO (and perhaps NARA, NTIS and other agencies) will maintain a publicly accessible GILS source that provides a comprehensive directory of all Core locators. When appropriate to their respective missions, Federal agencies may also develop and maintain additional interagency, topical locators that will also serve to enhance opportunities for sharing information resources. The following are examples of topics that might be the subject of additional interagency locators: economic indicators, trade information, spatial data, educational and training resources, disaster relief, health information, biodiversity and global change research. Such locators would be similar in function to the GILS Core, but would not necessarily use the GILS Core Elements format nor be focused solely on Federal agency holdings.

#### **4. Core Requirements**

##### **4.1 Functions**

GILS uses networking technology to provide a seamless facility that spans a wide variety of decentralized information sources. Within this range of sources, a subset will be Federal agency-based locators containing records that comply with the defined standards for GILS Core Elements. The GILS Core is defined as those locator records maintained by the U.S. Federal government, all of which comply with the defined GILS Core Element standards, and all of which are mutually accessible through interconnected electronic network facilities. Each information disseminating agency is responsible for compiling and maintaining their respective records in the GILS Core. Information services for access to GILS Core locator records will be maintained by Federal agencies without charge to the direct user.

The GILS Core is designed to satisfy Federal agency responsibilities to maintain an inventory of their electronic information dissemination products, as described in OMB Circular A-130. It should also be useful to agencies in improving agency responsiveness to FOIA requests. By including a record for each Federal information system holding publicly accessible data or information, the GILS Core thereby supports records management responsibilities of Federal agencies in reporting on agency information systems, codified in 44 U.S.C. Chapters 31 and 33. However, maintaining in GILS a reference to the availability of an information product does not itself satisfy all agency obligations under 44 U.S.C.

It is important to note that the vast majority of information sources accessible to GILS direct users would not be considered part of the GILS Core because they are not maintained by the U.S. Government, do not offer records in the format of the GILS Core Elements, are not on public networks, or are not offered free of charge. Many of these non-Core sources are locators nonetheless and will be very valuable for users in finding information. Also, other relevant sources of Federal information and Federal government information systems may be accessible to direct users of GILS. For example, various agencies and private sector information providers may develop products which contain GILS Core locator records. Indeed, such derivative and value-added products may often be the first point of access to Federal information resources.

The GILS Profile provides for the GILS locator records to be available in multiple forms, including Generic Record Syntax, United States Machine Readable Cataloging (USMARC), and Simple Unstructured Text Record Syntax (SUTRS). When using the Generic Record Syntax, the GILS locator elements can support representation in Hypertext Markup Language (HTML). (HTML is the format interpreted by the NCSA Mosaic client software when presenting World Wide Web objects, for example.) Provision has also been made in the GILS profile to support switching among navigation techniques, including use of a browsing mode as in gopher, or a searching mode as in Wide Area Information Servers (WAIS). The incorporation in GILS of Uniform Resource Identifiers (URIs) greatly simplifies electronic navigation among locators and other data systems available on interconnected networks.

## **4.2 Content**

The GILS Core will include records for all information locators that catalog other publicly accessible information resources at least partially funded by the Federal government, as well as for each of the Federal government information systems that include publicly accessible data or information. While GILS Core records can point to any kind of information source, they are especially designed for helping users navigate among a wide array of other locators of various forms.

It is not recommended that agencies use the precise format of the GILS Core locator records to describe all types of information resources. Rather, the agency should maintain inventory records

in a format appropriate to the primary user community being served. For example, the GILS Core Elements format would be a poor choice for describing each agency expert in particular technical areas, but it could well be used to describe the resource that contains a compilation of such descriptions. When such inventories are published, the originating agency should include a locator record that enables electronic linkage from and to the GILS Core locator.

The entire GILS Core is not likely to contain more than 100,000 locator records. In addition to locator records for information systems, it is estimated that the GILS Core will contain up to 1,000 locator records per Federal agency that is a major disseminator of public information. Agencies that are not major disseminators will typically have fewer records in their portion of the GILS Core, especially if the agency is relatively small. Where agencies maintain information inventories that have far more records, the agency is expected to aggregate related information resources in an locator record included in the GILS Core and to link the detailed inventory to GILS. Each GILS Core locator record is estimated to be less than 1,000 words in length. (Agency supplemental information, of course, may result in much larger locator records in some cases.)

### **4.3 Core Element Definitions**

Content definitions describe the GILS Core Elements required for users to determine the relevance of defined information resources to his or her need and to understand subsequent actions to obtain the information resources. These definitions identify relations among GILS Core Elements, and between GILS Core Elements and the USMARC format for bibliographic data. Terms used elsewhere and USMARC tags that appear to have similar content definitions are listed here as "Related Terms."

ANSI Z39.50 definitions for GILS Core Elements provide a structure and format for movement of the GILS elements between computer systems, such as in an on-line, local, or wide area networking environment. The Abstract Record Syntax and Basic Encoding Rules used to define GILS Core Elements are also suitable for movement of element contents between automated systems using digital media such as tape, diskette, or CD-ROM.

The GILS Profile offers preferred nomenclatures and templates of presentation formats for use in printed media as well as in electronic presentations. Although specified for human viewing in English, these are intended to also be extended to other languages. Separate templates may be appropriate for representing GILS Core Elements: online via Unstructured Text; online via HTML; online via a 24-line by 80-character computer terminal; and off-line in paper copy print.

#### 4.3.1. Mandatory Elements

**Title:** This mandatory element occurs once per locator record. It conveys the most significant aspects of the referenced resource and is intended for initial presentation to users independently of other elements. It should provide sufficient information to allow users to make an initial decision on likely relevance. It should convey the most significant information available, including the general topic area, as well as a specific reference to the subject.  
(Related Terms - USMARC 245\$a, heading, table of contents entry)

**Control Identifier:** This mandatory element occurs once per locator record. It is defined by the information provider and is used to distinguish this locator record from all other GILS Core entries. The control identifier should be distinguished with the record source agency acronym as provided in the U.S. Government Manual. (Related Terms - control number, system ID, URI)

**Abstract:** This mandatory element occurs once per locator record. It presents a narrative description of the information resource. This narrative should provide enough general information to allow the user to determine if the information resource has sufficient potential to warrant contacting the provider for further information. The abstract should not exceed 500 words in length. (Related Terms - USMARC 520, description)

**Purpose:** This mandatory element occurs once per locator record. It describes why the information resource is offered and identifies other programs, projects, campaigns, and legislative actions wholly or partially responsible for the establishment or continued delivery of this information resource. It may include the origin and lineage of the information resource, and related information resources. (Related Terms - USMARC 500, background, history)

**Originator:** This mandatory element occurs once per locator record. It identifies the information resource originator, named as in the U.S. Government Manual where applicable.  
(Related Terms - USMARC 710 with \$4org, creating organization)

**Access Constraints:** This mandatory element occurs once per locator record, although in some cases this element may contain the value "None." It describes any constraints or legal prerequisites for accessing the information resource or its component products or services. This includes any constraints applied to assure rights of privacy or intellectual property, and any other special restrictions or limitations on obtaining the information resource. Guidance on obtaining any users' manuals or other aids needed for the public to reasonably access the information resource must also be included here. (Related Terms - USMARC 506)

**Use Constraints:** This mandatory element occurs once per locator record, although in some cases this element may contain the value "None." It describes any constraints or legal prerequisites for using the information resource or its component products or services. This includes any constraints applied to assure rights of privacy or intellectual property and any other special restrictions or limitations on using the information resource. (Related Terms - USMARC 540)

**Availability:** This mandatory element occurs one or more times per locator record. It is a grouping of sub-elements that together describe how the information resource is made available.

**Distributor:** This mandatory sub-element occurs once per Availability element. It identifies the distributor by name. (Related Terms - USMARC 037)

**Resource Description:** This optional sub-element occurs no more than once per Availability element. It identifies the resource as it is known to the distributor. (Related Terms - USMARC 037)

**Order Process:** This mandatory sub-element occurs once per Availability element. It provides information on how to obtain the information resource from this distributor, including any fees associated with acquisition of the product or use of the service, order options (e.g., available in print or digital forms, PC or Macintosh versions), order methods, payment alternatives, and delivery methods. (Related Terms - USMARC 037)

**Technical Prerequisites:** This optional sub-element occurs no more than once per Availability element. It describes any technical prerequisites for use of the information resource as made available by this distributor. (Related Terms - USMARC 538)

**Available Spatial Reference:** This optional sub-element may occur multiple times per Availability element. When present, it provides the geographic reference for the information resource as made available by this distributor. (Formats are as given for the Spatial Reference element described below).

**Available Time Period:** This optional sub-element may occur multiple times per Availability element. It provides the time period reference for the information resource as made available by this distributor. (Time period formats are as given for the Time Period of Content element described below).

**Available Linkage:** This optional sub-element occurs no more than once per Availability element. It provides the information needed to contact an automated system made available by this distributor, expressed in a form that can be interpreted by a computer (i.e., URI). Available linkages are appropriate to reference other locators, facilitate electronic delivery of off-the-shelf information products, or guide the user to data systems that support analysis and synthesis of information. (Related Terms - USMARC 856, URI)

**Available Linkage Type:** This optional sub-element occurs if there is an Available Linkage described. It provides the data content type (i.e., MIME) for the referenced URI.

**Point of Contact for further information:** This mandatory element occurs once per locator record. It identifies an organization, and a person where appropriate, serving as the point of contact plus methods that may be used to make contact, such as telephone number, mail address, electronic mail address, fax number. (Related Terms - USMARC 856\$m for electronic resources, USMARC 500 for other than electronic resources)

**Record Source:** This mandatory element occurs once per locator record. It identifies the organization, as named in the U.S. Government Manual, that created or last modified or verified this locator record. (Related Terms - USMARC 040, responsible organization)

**Date Last Modified:** This mandatory element occurs once per locator record. It identifies the latest date on which this locator record was created, modified, or verified. (Related Terms - USMARC 008/00-05)

#### 4.3.2. Elements Mandatory for Information Systems

The GILS Core includes a locator record for each Federal information system holding publicly accessible data or information. The following two elements are optional for other GILS Core locator records.

**Agency Program:** This optional element occurs no more than once per locator record. It identifies the major agency program or mission supported by the system and should include a citation for any specific legislative authorities associated with this information resource. (Related Terms - USMARC 506\$e)

**Sources of Data:** This optional element occurs no more than once per locator record. It identifies the primary sources or providers of data to the system, whether within or outside the agency. (Related Terms - USMARC 537)

## 4.3.3. Optional Elements

**Controlled Vocabulary:** This optional element may occur multiple times per locator record. It is a grouping of sub-elements that together provide any controlled vocabulary used to describe the resource and the source of that controlled vocabulary.

**Index Terms - Controlled:** This sub-element occurs once per Controlled Vocabulary element. It is a grouping of descriptive terms drawn from a controlled vocabulary source to aid users in locating entries of potential interest. Each term is provided in the subordinate repeating field **Controlled Term**.  
(Related Terms - USMARC 650, keywords)

**Thesaurus:** This element occurs once per Controlled Vocabulary element. It provides the reference to a formally registered thesaurus or similar authoritative source of the controlled index terms. Notes on how to obtain electronic access to or copies of the referenced source should be provided, possibly through a Cross Reference to another locator record that more fully describes the standard and its potential application to locating GILS information.  
(Related Terms - USMARC 650\$2)

**Local Subject Index:** This optional element occurs no more than once per locator record. It is a grouping of descriptive terms to aid users in locating entries of potential interest, but the terms are not drawn from a formally registered controlled vocabulary source. Each term is provided in the repeating sub-element **Local Subject Term**. (Related Terms - USMARC 653, keywords)

**Methodology:** This optional element occurs no more than once per locator record. It identifies any specialized tools, techniques, or methodology used to produce this information resource. The validity, degree of reliability, and any known possibility of errors should also be described  
(Related Terms - USMARC 567, sensor, sampling, model)

**Spatial Reference:** This optional element occurs no more than once per locator record and provides the geographic reference for the information resource. Geographic names and coordinates can be used to define the bounds of coverage. Although described here informally, the spatial object constructs should be as defined in FIPS 173, "Spatial Data Transfer Standard."

**Bounding Rectangle:** This optional sub-element occurs no more than once within a Spatial Reference element. It provides the limits of coverage expressed by latitude and longitude values in the order: **western-most, eastern-most, northern-most, southern-most**.

**G-Polygon:** This optional sub-element may occur multiple times within a Spatial Reference element. It provides the actual outline of coverage, including voids, through two associated constructs. An **Outer G-Ring** represents the closed non-intersecting boundary of an interior area, and an **Exclusion G-Ring** represents the closed non-intersecting boundary of a void in an interior area.

**Geographic Name:** This optional sub-element may occur multiple times within a Spatial Reference element. It identifies significant areas and/or places within the coverage through two associated constructs: a **Geographic Keyword Name** and a **Geographic Keyword Type**. A preferred source of the names and types is the Geographic Names Information System.

**Coordinate Pair:** This optional sub-element may occur multiple times within a Spatial Reference element. It provides a representative location expressed by **latitude and longitude**.

**Time Period of Content:** This optional element may occur multiple times per locator record. It provides time frames associated with the information resource, in one of two forms:

**Time period - structured:** Time described using the USMARC prescribed structure. (Related Terms - USMARC 045)

**Time period - textual:** Time not described in the USMARC prescribed structure. (Related Terms - USMARC 500)



**Cross Reference:** This optional element may occur multiple times per locator record. Each instance is a grouping of sub-elements that together identify another locator record likely to be of interest. (Related Terms - USMARC 787)

**Cross Reference Title:** This optional sub-element occurs no more than once per Cross Reference element. It provides a human readable textual description of the cross reference.

**Cross Reference Linkage:** This optional sub-element occurs no more than once per Cross Reference element. It provides the machine readable information needed to perform the access. (Related Terms - URI)

**Cross Reference Type:** This optional sub-element occurs if there is a Cross Reference Linkage described. It provides the data content type (i.e., MIME) for the referenced URI.

**Original Control Identifier:** This optional element occurs no more than once per locator record. It is used by the record source agency to refer to another GILS locator record from which this locator record was derived. (Related Terms - control number, system ID, URI)

**Agency Supplemental Information:** This optional element occurs no more than once per locator record. Through this element, agencies may associate other descriptive information with the GILS Core locator record. (Related Terms - USMARC 500)

**Appendix 1: Glossary**

**agency** - any executive department, military department, government corporation, government controlled corporation, or other establishment in the executive branch of the United States Federal government, or any independent regulatory agency (OMB Circular A-130).

**ANSI Z39.50** - The "American National Standard Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection" is developed by the National Information Standards Organization (NISO), accredited to the American National Standards Institute (ANSI). ANSI Z39.50 complies with the Open Systems Interconnection (OSI) family of standards promulgated by the International Standards Organization (ISO), and is interoperable with the international standards for information search and retrieval, ISO 10162 and 10163. As of this writing, the currently approved version is ANSI Z39.50 Version 2.

**direct user** - a person or automated process that accesses GILS from networks using the GILS Profile and thereby having more flexibility to explore the full complement of available information. People who are direct users of GILS are assumed to be literate in English to at least the secondary school level, capable of using a personal computer, and aware of any constraints of their own hardware or software environment.

**dissemination** - the government initiated distribution of information to the public, excluding distribution limited to government employees or agency contractors or grantees, intra-agency or inter-agency use or sharing of government information, and responses to requests for agency records under the Freedom of Information Act (5 U.S.C. 552) or Privacy Act. Here, "disseminating information" is not distinguished from "providing access to information" (following OMB Circular A-130).

**government information** - information created, collected, processed, disseminated, or disposed of by or for the Federal government (OMB Circular A-130).

**Government Information Locator Service (GILS)** - a decentralized collection of locators and associated information services used by the public either directly or through intermediaries to find public information throughout the U.S. Federal government.

**GILS Core** - those sources maintained by the U.S. Federal government, all of which comply with the defined GILS Core Element standards and are mutually accessible through interconnected electronic network facilities without charge to the direct user. Although the GILS Core will be implemented initially on the Internet, it is intended to support broad interoperability.

**government publication** - information which is published as an individual document at government expense, or as required by law (OMB Circular A-130).

**information** - any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms (OMB Circular A-130).

**information product**- any book, paper, map, machine-readable material, audiovisual production, or other documentary material, regardless of physical form or characteristic (OMB Circular A-130).

**information resource** - includes both government information and information technology (OMB Circular A-130).

**information service** - considered equivalent to information product from the policy perspective of OMB Circular A-130, although agency locator records for services may differ from those for products.

**information system** - the organized collection, processing, maintenance, transmission, and dissemination of information in accordance with defined procedures, whether automated or manual (OMB Circular A-130).

**information technology** - the hardware and software operated by a Federal agency or by a contractor of a Federal agency or other organization that processes information on behalf of the Federal Government to accomplish a Federal function (OMB Circular A-130).

**intermediary or intermediate service** - an entity or service that makes some of the GILS information available but does not provide the full capabilities of a direct user.

**interoperability** - a condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems.

**locator** - an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in how to obtain the information.

**Open Systems Interconnection (OSI)** - a family of standards promulgated by the International Standards Organization (ISO) and adhering to a specific model that promotes interoperability.

**profile** - a set of implementor agreements providing guidance in applying a standard interoperably in a specific limited context.

**records management** - the planning, controlling, directing, organizing, training, promoting, and other managerial activities involved with respect to records creation, records maintenance and use, and records disposition in order to achieve adequate and proper documentation of the policies and transactions of the Federal government and effective and economical management of agency operations. (44 U.S.C. 2901(2))

**Uniform Resource Identifier (URI)** - A class of objects that defines a set of standards for the encoding of system independent resource location and identification information for the use of Internet information services. Examples of instantiations of this class include Uniform Resource Locators and Uniform Resource Names.

**USMARC** - USMARC is an implementation of ANSI/NISO Z39.2, the American National Standard for Bibliographic Information Interchange. The USMARC format documents contain the definitions and content designators for the fields that are to be carried in records structured according to Z39.2. GILS records in USMARC format contain fields defined in USMARC Format for Bibliographic Data and USMARC Format for Holdings Data. Both of these documents are published by the Library of Congress.

**Appendix 2: Extracts from Draft GILS Profile**

Note: In this Appendix, the author has extracted from a draft document entitled "Using Z39.50 In An Application For The Government Information Locator Service (GILS)." That document is being developed through a research project coordinated by Syracuse University and the United States Geological Survey, funded by the Federal Interagency Working Group on Data Management for Global Change. The complete draft document is available on the Internet via anonymous FTP (File Transfer Protocol) from 128.230.33.81 as /USGS/gils\_profile.txt (ASCII text format), or by mail from William E. Moen, Syracuse University, School of Information Studies, 4-206 Center for Science and Technology, Syracuse, NY 13244-4100. Telephone 315-443 -4508. Comments can be submitted via electronic mail to: [wemoen@mailbox.syr.edu](mailto:wemoen@mailbox.syr.edu).

**INTRODUCTION**

This document describes an ongoing research effort to develop a profile for the use of ANSI/NISO Z39.50, The American National Standard Information Retrieval Application Service Definition and Protocol Specification for Library Applications (National Information Standards Organization, 1992), in the proposed Government Information Locator Service (GILS)...ANSI/NISO Z39.50 is an American National Standard developed and approved by the National Information Standards Organization (NISO)...The purpose of Z39.50 is to allow one computer operating in a client mode to perform information retrieval queries against another computer acting as an information server...The standard is an applications-layer protocol within the Open Systems Interconnection (OSI) reference model...Z39.50 is parallel to two international standards: ISO 10162: 1993 Information and documentation -- Search and Retrieve Application Service Definition; and ISO 10163-1: 1993 Information and documentation -- Search and Retrieve Application Protocol Specification...

A profile is "a set of one or more base standards, and where applicable, the identification of chosen classes, subsets, options and parameters of those bases standards, necessary for accomplishing a particular function...Profiles are also referred to as "functional standards," "implementation agreements," or "specifications."...The research team broadened this definition for the GILS Profile to include not only the specifications for Z39.50 in the application but also other aspects of the implementations for a GILS conformant locator that are beyond the scope of the base standard (i.e., Z39.50)...

The GILS Profile provides the specifications for the overall GILS application relating to the GILS Core and will completely specify the use of Z39.50 in this application...This first version of the GILS Profile is focused on the requirements for a GILS server. GILS clients will be able to interconnect with any GILS server, and these clients will behave in a manner which allows interworking with the GILS server. Clients that support Z39.50 but do not implement the GILS profile will be able to access GILS records with less than full GILS functionality...

**ASSUMPTIONS AND AGREEMENTS ABOUT GILS**

The GILS is understood to be an agency-based, Internet- accessible locator service. "Direct users" ...will connect to the GILS via the Internet using Z39.50 clients and servers to find information about a wide range of Federal information resources.

Agencies will develop and maintain Locators. Locators are machine-readable databases that contain Locator Records describing Federal information resources...The GILS Profile does not specify the base technology (e.g., a database management system) that an agency uses to mount its Locator database nor does it specify internal storage of records in the database...

A GILS Locator accessed using Z39.50 in the Internet environment acts primarily as a pointer to information resources. Some of these information resources pointed to by GILS Locator Records, as well as the GILS Locator, may be available electronically through other communications protocols including the common Internet protocols that facilitate electronic information transfer such as remote login (Telnet), File Transfer Protocol (FTP), and electronic mail (SMTP/MIME). The use of these protocols or other communications paths is outside the scope of this project and of the GILS Profile...

The GILS Core...contain[s] individual Locator Records, structured with a standardized set of data elements (i.e., GILS Core Elements), that provide summary descriptions of Federal information resources. These Locators (i.e., machine-readable databases) are themselves Federal information resources and can be described by Locator Records...

Direct users must have prior knowledge of at least one of the GILS Locators and its network address, and must be able to access it to enter the GILS. Upon entry, however, users may follow links provided in the Locator Records to navigate through Locators existing on a number of servers. The semantics of the Locator Records coupled with a client that understands these semantics and building upon the ability of the Z39.50 protocol to provide a uniform interface to multiple autonomously managed servers combine to provide the user with the impression of seamless navigation among these distributed servers. The semantics of the Locator Records allow elimination of duplicate records, further fostering the impression of a single system built out of autonomous, distributed servers.

Each of the GILS Core Locators can be represented by a Locator Record in other GILS Core Locators. Some GILS Core Locators will include references to all of the GILS Core Locators, and these might be regarded as a kind of "directory of directories." However, GILS itself does not assign any hierarchical status to specific locators. Rather, the structure and content of the GILS Locator Records enable, for example, the aggregation of Locator Records in "directories" that could be offered by one or more Federal agencies or other organizations.

Once connected to the GILS, users may navigate through single or multiple Locators. GILS servers will support searching (i.e., accept a search query and return a result set or diagnostic messages) and may support browsing (i.e., accept a well-known search query and return a list of Locator Records in brief display format). GILS servers must be able to return all elements of Locator Records, or combinations of those elements, that contain non-zero length data.

A Locator Record consists of a number of data elements that identify and describe an information resource...Several data elements can be included in Locator Records to facilitate GILS navigation and electronic network- based access to information...

Users will be able to search a Locator as a means of finding out how to acquire or access the information resource described by one or more Locator Records. GILS servers may support a variety of search strategies...A user's search specification is received by a Locator (GILS server) using the Search Facilities of Z39.50. The searchable elements of the Locator Records are called Attributes...The exact manner by which the user constructs the query is an interface issue and not specified by the profile, but the user must be able to specify searches with each of the required Attributes...

After a GILS server completes a search, it produces a result set and makes that available to a client. The GILS server provides the client the contents of selected records from the result set using the Retrieval Service of Z39.50. The GILS server must respond to requests that records be served up in one of the three Element Sets...specified by the GILS Profile. The exact manner in which a result set is presented to the user is also an interface issue and not within the scope of the profile.

A GILS Locator may provide a structure for browsing that is comprised of a chain of Locator Records traversed through pointers specified in the GILS Core Element CROSS REFERENCE. The CROSS REFERENCE is a repeating element. Each occurrence contains a item pointer in the form of a Uniform Resource Locator (URL), the title of the item, and a content type to identify it. Each referenced item may be a Locator Record on the same Locator or on another Locator.

To support browsing, an agency may include among the Locator Records on a GILS server one with a zero-length value for its CONTROL IDENTIFIER. A GILS server will include this record in result set in response to a well-known search query... This allows users to browse a Locator when or if they have no other starting point. If, in response to this well-known search query, the result set is empty, this particular Locator does not contain such a record...

### Attribute Sets

The profile specifies a GILS Attribute Set that...consists of all Bib-1 Attributes and other Use Attributes that are defined for GILS elements that cannot be mapped to Bib-1 Use Attributes. For any additional Use Attributes that cannot be mapped to Bib-1 Use Attributes, these will be numbered in sequence beginning at 2000 and ending at 2999. These are well-known attributes and will correspond in name and semantics to the elements in the GILS Schema. Since the use of Generic Record Syntax (GRS) allows the creation of additional, agency- or originator- defined string-tagged elements, the GILS Attribute Set allows these not-well-known elements to be identified as attributes with tags numbered above 3000...

### Diagnostic Messages

The standard provides a list of diagnostic messages that can be exchanged in the course of an association between an origin (client) and target (server). The GILS application will use Diagnostic Set Bib-1.

### Record Syntaxes

Record syntaxes provide for the transfer of database records between a target (server) and an origin (client) in acceptable form for processing. The profile requires servers to support the following three record syntaxes:

- USMARC -- an implementation of ANSI/NISO Z39.2 and maintained by the Library of Congress
- Generic Record Syntax (GRS) -- defined in Z39.50
- Unstructured Text (SUTRS) -- defined in Z39.50...

The Generic Record Syntax is a general-purpose format for packaging records of varying complexity with potentially arbitrary data in individual fields. For mainly-text records like GILS records, GRS is simple and efficient. USMARC is a format used by many bibliographic systems. These systems are likely to be important users of GILS...Unstructured Text (SUTRS) provides a bare-minimum operating capability. SUTRS records consist of a single text field formatted by the target system (server). GILS targets (servers) will use the Preferred Presentation Format to format Locator Records for Unstructured Text transmission...In the cases when the server transmits records using SUTRS or USMARC record syntaxes, it will alert the user that information will be lost and the data transformations will not necessarily be reversible.



## Z39.50 SPECIFICATIONS FOR THE GILS APPLICATION

The GILS Profile details a range of facilities and services available in Z39.50, describes an Attribute Set for searching and three Element Sets by which the server presents some or all the elements of the Locator Records, and prescribes the Record Syntaxes to be supported by GILS servers for the transfer of Locator Records...The terminology and concepts presented in this section are specific to this standard. Readers should consult the complete standard for further information and reference. For example, the standard uses the words "origin" and "target," rather than "client" and "server."...

GILS Origins (clients) and Targets (servers) support Z39.50, Version 2...

Facilities

GILS Z39.50 Origins (clients) and Targets (servers) must support the following Version 2 Facilities and Services for information retrieval for operation in the Internet environment:

FACILITYSERVICE

Init Facility -- allows an origin (client) to propose values for initialization parameters.

Init Service

Search Facility -- enables an origin system (client) to query a database at a target system (server), and to receive information about the results of query.

Search Service

Retrieval Facility -- enables the origin (client) to retrieve records according to position within a result set maintained by the target (server).

Present Service

Termination Facility -- allows the origin (client) or target (server) to initiate abrupt termination or graceful termination of a connection.

Mapped to TCP ABORT  
or TCP CLOSE

Standard Z39.50 Init Service negotiation procedures control the use of all services. No additional services are required for conformance to the GILS Profile. Other Z39.50 services, however, may be provided optionally by target systems (servers) and used by origins (clients).

Search Service Parameters

The GILS application will support Z39.50 Type 1 queries which are Reverse Polish Notation (RPN) queries.

**Preferred Presentation Format**

The profile recommends a preferred presentation format for SUTRS records. For the SUTRS records, formatting instructions for a preferred presentation format is a concern of the server. The preferred presentation format is not intended to provide a structure for SUTRS records that enables parsing. In addition, the profile will suggest, but not prescribe, display formats for GRS and MARC records.

**Schema**

The GILS Profile specifies a GILS Schema that will be a registered object ...The GILS Schema will use Schema-1 elements and define additional elements as necessary. The profile will specify tag types to identify which are Schema-1 elements (Tag Type = 1) and which are additionally defined elements of the GILS Schema (Tag Type = 2)...Schema elements can be nested and the tagging notation will reflect the nesting...

Any well-known element will be assigned a numeric tag. GRS provides a flexibility that allows additional information to be identified by elements, and these agency-defined elements will use string-tags for identification. The string-tags in Core Elements that are not subfielded are agency defined and are not well-known, and thus not defined in the GILS Schema

**Element Sets**

The profile specifies three Element Sets that GILS servers must support. Each Element Set consisting of [the following] elements from the GILS Schema:

- B -- contains at least Title, Control Identifier, Originator, Date Last Modified, and Local Control Number
- G -- contains all B Element Set elements and Cross References
- F -- contains all elements available in the record.

**CONCLUSION**

...The goal of this research project is to ensure that the GILS Profile is implementable and usable, and that implementations based on the Profile can interoperate and interwork. Achieving this goal will serve the larger goals of the Government Information Locator Service by providing a standards-based, decentralized, network-accessible service through which the public will be able to identify and locate Federal information resources. In addition, the GILS Profile provides the means by which various implementors using a variety of computer platforms (clients and servers) can develop products usable by Federal agencies and the public.



REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE August 1994	3. REPORT TYPE AND DATES COVERED Technical Memorandum	
4. TITLE AND SUBTITLE NASA STI Program Coordinating Council, Twelfth Meeting: Standards		5. FUNDING NUMBERS	
6. AUTHOR(S)			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Headquarters		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546		10. SPONSORING/MONITORING AGENCY REPORT NUMBER  TM 109925	
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified - Unlimited Subject Category 82		12b. DISTRIBUTION CODE	
13. ABSTRACT (maximum 200 words)  The theme of this NASA Scientific and Technical Information Program Coordinating Council Meeting was standards and their formation and application. Topics covered included scientific and technical information architecture, Open Systems Interconnection Transmission Control Protocol/Internet Protocol, Machine-Readable Cataloging open system environment procurement, and the Government Information Locator Service.			
14. SUBJECT TERMS NASA programs, information systems, procurement, computer networks, standards, specifications, computer systems programs, libraries		15. NUMBER OF PAGES 213	
		16. PRICE CODE HC-A10/MF-A03	
17. SECURITY CLASSIFICATION OF REPORT UNCLASS	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASS	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASS	20. LIMITATION OF ABSTRACT UNLIMITED



**POSTMASTER**  
*Address Correction Requested*  
*(Sections 137 and 159 Post Manual)*

**National Aeronautics and  
Space Administration  
Code JTT  
Washington, DC 20546-0001**

*Official Business*  
*Penalty for Private Use, \$300*

**BULK RATE**  
**POSTAGE & FEES PAID**  
**NASA**  
**PERMIT No. G-27**